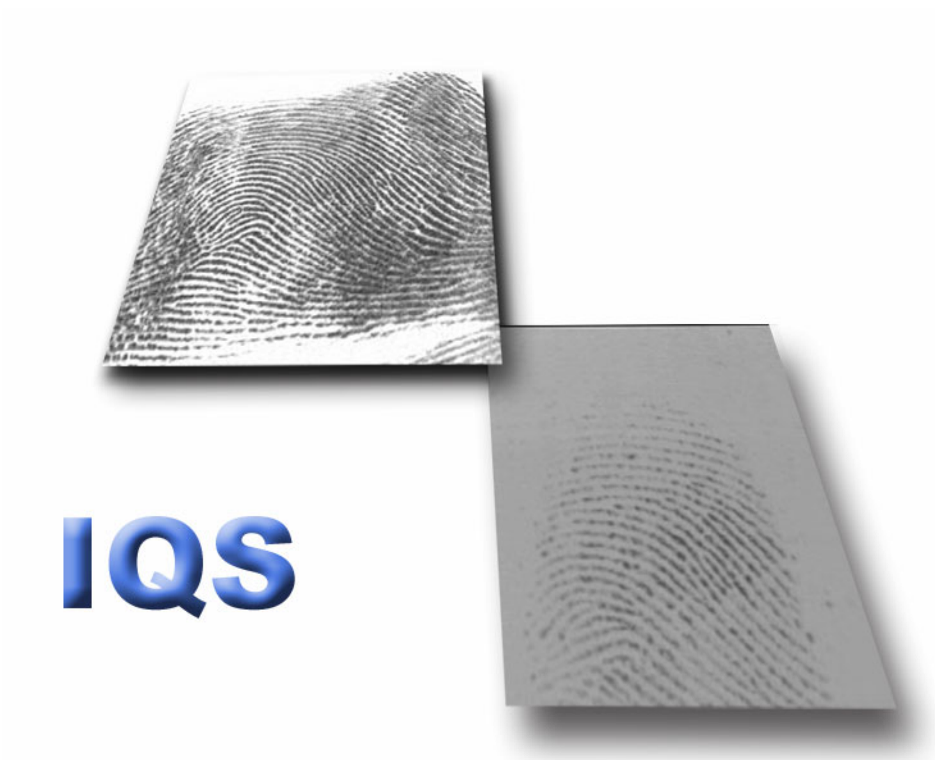
Implications of the IDENT/IAFIS Image Quality Study for Visa Fingerprint Processing



October 31, 2002

R. Austin Hicklin

Christopher L. Reedy, PhD

MTS™ Mitretek Systems

Abstract

The purpose of this document is to review the findings of the Image Quality Study (IQS), a fingerprint performance study conducted by Mitretek Systems in 2000, and to state the implications of this study and subsequent analyses for large-scale visa fingerprint processing.

The National Institute of Standards and Technology (NIST) needs to determine the accuracy of the Federal Bureau of Investigation's (FBI's) Integrated Automated Fingerprint Identification System (IAFIS) for use in visa processing. Some of the necessary testing was performed as part of the IQS. The IQS was developed to determine how the FBI's IAFIS—with fingerprints of more than 40 million individuals—would perform when searched with flat impressions of two index fingers collected by the Immigration and Naturalization Service's IDENT system. This study was expanded to predict IAFIS performance when searched with an arbitrary number of up to ten flat fingerprint impressions.

The results of this report differ somewhat from those of the IQS Report since the IQS Report had a specific focus that was different from the visa fingerprint processing study. In addition, some of the results here are new. Since completion of the IQS, more extensive data analyses have been conducted, and new information has come to light.

The following are the key findings and recommendations of this report:

- Slap fingerprints are appropriate for use in large-scale identification systems. This is the optimal compromise between matcher performance and operational constraints. Use of slaps will improve system performance and reduce processing requirements when searching databases larger than 10 million subjects.
- Two-finger searches of IDENT-quality fingerprints cannot achieve adequate performance against the existing IAFIS without a dramatic increase in processing resources.
- Search fingerprint Image Quality Metrics alone are an imperfect predictor of search performance. However, poor search fingerprint quality is an effective predictor of search failure.
- Large identification systems should be multimodal, incorporating demographic, facial, and possibly other biometric data. The impact of errors arising from reliance on a single biometric can be largely overcome by incorporating alternative identifiers.
- A research program for ongoing analysis and comparison of emerging AFIS technology is needed. Investigations should be conducted to determine the availability of new or improved algorithms, the possibility of improving existing algorithms, and the potential impacts of each.
- Representative test data sets need to be collected for target search populations. Testing determined that female fingerprints are significantly lower quality than male prints. Test sets representing children and the elderly are needed.
- Policies and procedures to maintain operational quality need to be developed, including designing systems to measure ongoing operational performance.

Executive Summary

Introduction

The purpose of this document is to review the findings of the Image Quality Study (IQS), a fingerprint performance study conducted by Mitretek Systems in 2000, and to state the implications of this study and subsequent analyses for large-scale visa fingerprint processing.

The National Institute of Standards and Technology (NIST), in conjunction with the Attorney General and the Secretary of State, is required to submit a report to Congress assessing the actions and considerations needed to achieve implementation of a system using biometric identifiers. It is also tasked with developing associated standards for verifying the identity of individuals entering and exiting the United States and identifying individuals that are applying for entry into the United States. The proposed system requires use of the Federal Bureau of Investigation's (FBI's) Integrated Automated Fingerprint Identification System (IAFIS) for fingerprint-based criminal background checks. NIST is using the Algorithm Test Bed (ATB) to model IAFIS performance. The ATB, which was used by Lockheed Martin to design and test the algorithms and throughput performance of IAFIS, uses the same software and hardware matchers as IAFIS but on a smaller scale. A copy of the ATB is being configured by Lockheed Martin for NIST use.

NIST needs to determine the accuracy of IAFIS for use in visa processing. Some of the necessary testing was performed as part of the IQS. The IQS used the ATB to conduct a variety of tests using Immigration and Naturalization Service (INS) fingerprint data from August through December 1990. After the IQS, a small-scale preliminary analysis of the effectiveness of slap fingerprints was conducted on the ATB using FBI civil fingerprint data. Further analyses of the data collected in the IQS and slaps studies have yielded some new results, which are also reported in this document.

Purpose of the IQS

In 2000, the Department of Justice (DOJ) was developing a strategy to integrate the INS's IDENT system with IAFIS. Mitretek supported that activity by conducting an Engineering/System Development Study (E/SDS) to identify the requirements and architecture for the integrated system. One approach considered for the integrated IDENT/IAFIS system was to capture the two-finger INS data and search this data against rolled fingerprints in the IAFIS Criminal Master File (CMF). The quality and characteristics of the search and file fingerprints determine the hardware resource requirements and performance of an Automated Fingerprint Identification System (AFIS). IAFIS performance—when searched with rolled ten-prints—is well understood. However, the FBI had little experience searching flat two-prints against IAFIS.

The purpose of the IQS Study was to determine objectively how the FBI's IAFIS—with more than 40 million subjects in the CMF—would perform when searched with flat impressions of two index fingers. This study was expanded to project IAFIS performance when searched with an arbitrary number of up to ten flat fingerprint impressions. Key performance measures of interest were reliability, selectivity, and filter rate. An additional

aim of the IQS was to establish an image quality metric baseline, which would be useful when testing and monitoring the performance of the eventual integrated IDENT/IAFIS system.

Changes since the IQS

The results of this report differ somewhat from those in the IQS since the IQS had a specific focus that was different from the visa fingerprint processing study. In addition, some of the results presented in this report are new. Since the IQS's completion in early December 2000, more extensive analyses of the data have been conducted, and new information has come to light. These changes include the following:

- **Slap Fingerprint Analyses**. Immediately after the IQS was completed, Mitretek conducted a small number of tests on the ATB using slap fingerprint data obtained from the FBI. A short informal paper reported on the accuracy of commercial segmentation software from Aware Corporation, and the matcher performance of the segmented fingerprints.
- **New Analysis of IQS Data**. The IQS schedule made it impossible to process all of the data that was collected. Some new results have emerged following further analysis of the same data. These results are reported in this document.

Issues and Limitations

The implications of the IQS for visa processing are limited by several issues.

- ***IQS estimates should be used cautiously to estimate performance with populations or systems that differ significantly from those studied.***

  The IQS provided an accurate estimate of how two-finger flat data with characteristics specific to INS IDENT subjects would perform against the IAFIS CMF. However, IQS findings may be limited if a different population is considered or if different matcher algorithms—including a retuned IAFIS—are used.

- ***Current IAFIS performance may be better than indicated by the IQS.***

  As a result of the FBI's Technology Refreshment Program (TRP), the algorithms used in IAFIS are known to have improved since the IQS, most notably in the area of pattern classification.

- ***Slap performance has not been adequately tested.***

  The slap fingerprint study conducted after the IQS was a limited analysis, based on a small data set that may not have been representative of FBI Civil fingerprint submissions. Results from the slap fingerprint analysis should be regarded as preliminary; they will be replaced as results from more complete studies become available.

Findings

- ***Four or more flat fingerprints—and preferably six or more—should be used when searching databases larger than 10 million subjects.***

  The IQS showed that IAFIS could not meet its accuracy requirements using ten-print algorithms with two-finger searches of IDENT-quality data. IAFIS could meet its accuracy requirements for two-finger searches using latent algorithms, but only at significantly increased processing cost. Searching with four or more fingers will result in acceptable accuracy.

- ***Additional fingerprints significantly reduce processing requirements for searching large databases.***

  Using more fingers significantly improves processor performance. This improvement derives from the use of fingerprint classification indexing to reduce the number of candidates for each search. For each pair of fingers included in the search prints, the partitioning algorithm is able to cut the number of potential candidates approximately in half, which in turn halves processor requirements.

- ***The existing IAFIS algorithms could be reengineered to form a basis for improved flat fingerprint processing.***

  Portions of the current IAFIS ten-print and latent algorithms could be combined to produce a system with flat fingerprint performance superior to the existing ten-print system and processing requirements significantly lower that the existing latent system.

- ***Female fingerprints are poorer quality than male fingerprints.***

  A greater proportion of female fingerprints are poor or very poor quality. On average, matching female fingerprints will require about 150 percent of the processing needed to match male fingerprints. Clearly, performance and throughput will be engineering challenges for systems with large female populations.

- ***Improvements in operational fingerprint quality will improve search accuracy.***

  The findings in the IQS are specific to the IDENT-quality data that was provided for this test. The preliminary slap fingerprint results indicate that just using the two index fingers from slap images would have substantially better performance accuracy than the two-finger IDENT data. This suggests that improvements in the operational quality of the data (such as by using more expensive fingerprint scanners) would improve performance accuracy over the IQS results.

- ***Operational fingerprint data will produce failure-to-enroll (FTE) errors.***

  Current IAFIS operations reject about 2.5 percent of civil submissions (rolled fingerprints) due to poor fingerprint quality. The quality of approximately 2 percent of INS IDENT flat fingerprints is so poor that it renders them virtually impossible to match using current IAFIS technology, and an additional 3 percent would be very unlikely to match. Slap fingerprints must be segmented into separate images; this

process has an associated segmentation error rate that may result in FTE in a small percentage of cases.

- ***Search fingerprint quality alone is an imperfect predictor of search performance.***

  Image quality of the search fingerprint is only one of a number of factors that determine the accuracy of fingerprint matching.

- ***Poor search fingerprint quality is an effective predictor of search failure.***

  Fingerprints with poor quality are very unlikely to match. Effective minimum quality thresholds can be established using one or more IQMs.

- ***The methods by which sample data sets are collected can bias them so strongly that they are unusable for testing.***

  Great care must be taken in collection of the data sets used for testing. In particular, mated data sets should never be selected by using an AFIS; this process essentially filters out all of the hard-to-match fingerprints.

Recommendations

- ***Slap fingerprints are appropriate for use in large-scale identification systems.***

  The significant improvement in accuracy and processing requirements as the number of search prints increases suggests that the use of slaps is the optimal compromise between matcher performance and operational constraints. The use of slaps offers operational improvements over the use of rolled fingerprints, since collecting slap fingerprints is a rapid process that does not require the same degree of operator training and “manhandling” of the subject. Operationally, collecting slaps and flat fingerprints is very similar. The use of slaps offers improvements in performance accuracy and efficiency over the use of flats.

- ***Large identification systems should be multimodal, incorporating demographic, facial, and possibly other biometric data.***

  The impact of errors arising from reliance on a single biometric can be largely overcome by incorporating alternative identifiers. An additional biometric would be particularly useful in processing subjects with poor quality fingerprints.

- ***Initiate a research program for ongoing analysis and comparison of emerging AFIS technology. Investigate the availability of new or improved algorithms, the possibility of improving existing algorithms, and the potential impacts of each.***

  There are a number of different areas of research that should result in improved identification system performance.

- ***Collect representative test data sets for target search populations.***

  Current test data sets were drawn largely from criminal populations and may not be representative of visa applicants. Test sets representing children and the elderly are particularly needed.

- ***Develop and standardize policies and procedures to maintain operational quality.***

  Systems should be designed so that the equipment and operators are capable of collecting fingerprints of adequate quality. In addition, ongoing measures (such as sampling or random tests) should be implemented to verify that the equipment and operators are in fact delivering fingerprints of adequate quality.

- ***Design systems to measure ongoing operational performance.***

  Operational quality policies and procedures should be implemented in identification systems. Without auditing or sampling to determine operational error rates, there is no means of determining ongoing system effectiveness. Lights-out systems should be instrumented to provide for audits. Template-only systems—those that do not store human-verifiable images that can be audited—have unknown operational performance.

Acknowledgments

The authors would like to acknowledge Donald D'Amato, Rajiv Khanna, George Kiebuzinski, Lawrence Nadel, and John Splain of Mitretek Systems who were the authors of the original IQS report and provided assistance and review for this revision. In addition, the authors also would like to thank the following individuals for their assistance with this document: Larry Pantzer, Linda Nichols, John King, and Nirav Desai. The following acknowledgements are quoted from the original IQS Report:

> *The authors would like to acknowledge Henry Culpepper, Art Forman, Jim O'Sullivan, and Don Ziesig of Lockheed Martin Corporation (LMC) for their contributions in developing the Algorithm Test Bed (ATB) and their Help Desk support during the use of the test bed.*
>
> *The authors also would like to thank the following individuals for their specific oversight and support of this study: Frank Boyle of the U.S. Department of Justice (DOJ), Justice Management Division (JMD); John Werner and Jeff Bowles of the Federal Bureau of Investigation (FBI), Criminal Justice Information Services (CJIS) Division; and Brad Wing of the U.S. Immigration and Naturalization Service (INS).*

Table of Contents

List of Figures

List of Tables

Section 1: Introduction

1.1 Purpose

The purpose of this document is to review the findings of the Image Quality Study (IQS), a fingerprint performance study conducted by Mitretek Systems in 2000, and of subsequent analyses and to state the implications of these studies for large-scale visa fingerprint processing.

Section 303 of the “Enhanced Border Security and Visa Entry Reform Act of 2002”, H.R. 3525, calls for the Attorney General and the Secretary of State to establish biometric identifier standards to be used for visas and other travel and entry documents. These standards are to be chosen from among those biometric identifiers recognized by domestic and international standards organizations. The National Institute of Standards and Technology (NIST), acting jointly with the Attorney General and the Secretary of State, is required to submit a report to Congress assessing the actions and considerations needed to implement a system using biometric identifiers and associated standards.

NIST is considering using fingerprints on travel and entry documents. NIST is testing the use of fingerprints and other biometrics both to verify the identity of individuals entering and exiting the United States and to identify individuals applying for entry into the United States.

The proposed system for visa screening uses the Federal Bureau of Investigation’s (FBI’s) Integrated Automated Fingerprint Identification System (IAFIS) system for fingerprint-based criminal background checks as part of the initial application process. Since IAFIS is an operational production system, it is impractical to test concurrently with production use. As an alternative to the use of the production IAFIS system for testing, NIST is using the Algorithm Test Bed (ATB) to model IAFIS performance. This system was used by the Lockheed Martin Corporation (LMC) to design and test the AFIS algorithms and throughput performance of IAFIS. The ATB uses the same software and hardware matchers as IAFIS on a smaller scale. Lockheed Martin is configuring a copy of the ATB for NIST use. As of the writing of this document, the ATB at NIST is not yet available to perform tests.

NIST needs to determine the accuracy of IAFIS for use in visa processing. Some of the necessary testing has already been performed as part of the IQS. The IQS used the ATB to conduct a variety of tests using INS fingerprint data from August 1990 to December 1990. After the IQS, a small-scale preliminary analysis of the effectiveness of slap fingerprints was conducted on the ATB, using FBI civil fingerprint data. Further analyses of the data collected in the IQS and slaps studies have yielded some new results, which are reported in this document.

1.2 The Image Quality Study (IQS)

1.2.1 Background: IAFIS and IDENT

IAFIS (Integrated Automated Fingerprint Identification System)

IAFIS is the FBI's automated ten-print and latent fingerprint identification system and criminal history file. The Automated Fingerprint Identification System (AFIS) segment of IAFIS is responsible for searching submitted fingerprints against the digitized fingerprints maintained by IAFIS. Within AFIS, the FBI's Criminal Master File (CMF) is composed of rolled ten-print records for more than 40 million arrestees. Except in the case of latent fingerprint searches, criminal and civil searches of the FBI's CMF are performed with rolled ten-print images. IAFIS was intended to be primarily a ten-print identification system, and special efforts were made to optimize IAFIS performance for rolled ten-print search data. The IQS study focused on the performance impact on IAFIS of searching INS fingerprints. The AFIS segment of IAFIS was developed by Lockheed Martin Corporation (LMC), with Sagem Morpho and CALSPAN as major subcontractors.

IDENT (INS's Automated Biometric Identification System)

IDENT, an INS automated biometric identification system used to monitor illegal border crossing activity, was designed to identify the recidivists among illegal border crossers for possible criminal prosecution. At border crossings (ports of entry) and border patrol stations, INS agents capture *flat* images of individuals' right and left index fingers to check the identity and criminal background of aliens attempting to enter the United States. The index finger images are first searched against the Lookout database, a rolled ten-print database of approximately 300,000 individuals with active 'wants.' Assuming a mate is not found, the two-print is then searched against the Recidivist database, a flat two-index-finger database of approximately 2.8 million records of individuals who have been caught previously attempting to cross the border illegally. The IDENT matcher was developed by Cogent Systems.

1.2.2 Purpose of the IQS Study

In 2000, the Department of Justice's (DOJ's) Justice Management Division (JMD) was developing a strategy to integrate IDENT with IAFIS. In support of that activity, Mitretek conducted an Engineering/System Development Study (E/SDS) to identify requirements and architecture for the integrated system. One of the E/SDS goals was to develop a strategy to integrate IDENT and IAFIS effectively while minimizing changes to either system. One approach considered for the integrated IDENT/IAFIS system was to capture the two-finger INS data and search this data against rolled fingerprints in the IAFIS CMF. The quality and characteristics of the search and file fingerprints determine the hardware resource requirements and performance of an AFIS. IAFIS performance, when searched with rolled ten-prints, is well understood. However, the FBI had little experience searching flat two-prints against IAFIS.

The purpose of the IQS Study was to determine objectively how the FBI's IAFIS, with more than 40 million subjects in the CMF, would perform when searched with flat impressions of two index fingers. This study was expanded to predict IAFIS performance

when searched with an arbitrary number of up to ten flat fingerprint impressions. Key performance measures of interest were reliability, selectivity, and filter rate.[1] The results obtained were expressed in a fashion suitable for use by E/SDS engineers to make performance, cost, and operational tradeoffs and to specify candidate system architectures for an IDENT/IAFIS system. An additional aim of the IQS was to establish an image-quality metric baseline, which would be useful when testing and monitoring the performance of the eventual integrated IDENT/IAFIS system.

It was beyond the scope of the IQS to identify improvements in the live-scan fingerprint capture process that might improve image quality, and thus improve the performance associated with searching flat fingerprints against IAFIS.

1.2.3 IQS Findings

The IQS study identified a number of factors that control flat to rolled fingerprint matching performance:

- Number of Fingers
- Correspondence between Search and File images
  - Overlapping areas
  - Lack of mutual distortion
- Quality of *both* Search and File images
  - Quality of ridge detail
  - Number of features
  - Size of image

The quality of the fingerprints used is critical, particularly if either the search or file print is of poor quality. The quality of approximately 2 percent of INS IDENT flat fingerprints is so poor that it renders them virtually impossible to match using current IAFIS technology, and an additional 3 percent would be very unlikely to match. Current IAFIS operations have a reject rate due to poor image quality of 0.5 percent for criminal search data and about 2.5 percent for civil search data. The number of searches that cannot be matched due to poor quality can be reduced by using more fingers or by improving the quality of the capture process.

This study concluded that the IAFIS ten-print algorithm suite cannot meet the IDENT/IAFIS reliability and selectivity requirements for a two-finger search. Using four or more (preferably six or more) fingers with the IAFIS ten-print algorithm suite is likely to produce results at the desired performance level, but it would require improvements in IAFIS capacity and workflow management. The use of more fingers not only increases system accuracy, it also dramatically reduces the size and cost of the necessary hardware; each additional pair of fingers (except the little fingers) used in a search approximately halves the AFIS processing requirements.

[1] Key concepts are defined and explained in Section 2.1, “Key Concepts and Terminology.”

1.3 Changes since IQS

The results reported in this report differ somewhat from those in the IQS since the IQS had a more specific focus: the performance of IDENT-quality two-finger flat fingerprints. In addition, some of the results presented in this report are new. Since the IQS's completion in early December 2000, more extensive analyses of the data have been conducted, and new information has come to light.

- ***Slap Fingerprint Analyses***

  Immediately after the IQS was completed, Mitretek conducted a small number of tests on the ATB using slap fingerprint data obtained from the FBI. A short informal paper[2] reported on the accuracy of Beta-release segmentation software from Aware Corporation, and the matcher performance of the segmented fingerprints.

  Segmentation software improved dramatically after the slaps study was completed. In February 2002, the same data set was segmented using a later commercial release of the Aware software with substantially better accuracy. An appendix to the original slaps study document was added to report the February 2002 results.

  The results from the slaps study are instructive but should be treated as preliminary; they will be modified as results from more complete studies become available.

- ***New Analysis of IQS Data***

  The schedule of the IQS made it impossible to process all of the raw data that was collected. Some new results have emerged based on further analysis of the same data. These results are reported in this document.

[2] Austin Hicklin, "Preliminary Analysis of Slap Fingerprint Performance," January 26, 2001.

1.4 Issues and Limitations

The implications of the IQS for visa processing are limited by several issues.

- ***IQS estimates should be used cautiously to estimate performance with populations or systems that differ significantly from those studied.***

  The IQS provided an accurate estimate of how two-finger flat data with characteristics specific to INS IDENT subjects would perform against the IAFIS CMF. However, IQS findings may be limited under the following conditions:

  - Search data has different population or operational characteristics
  - Different matcher algorithms are used
  - IAFIS is retuned for a new purpose

- ***Current IAFIS performance may be better than indicated by the IQS.***

  As a result of the FBI's Technology Refreshment Program (TRP), the algorithms used in IAFIS are known to have improved since the IQS, most notably in the area of pattern classification.

- ***Slap performance has not been adequately tested.***

  The slaps study conducted after the IQS was a limited analysis, based on a small data set that may not have been representative of FBI Civil fingerprint submissions. Results from the slaps analysis should be regarded as preliminary, to be replaced as results from more complete studies become available.

Section 2: Study Overview

This section provides a general background of the study, including key concepts, the overall approach to the study, and a description of the ATB and data sets used in the study.

2.1 Key Concepts and Terminology

2.1.1 Identification and Verification

Identification is a term used to describe the process of matching a biometric record from a single subject against an entire database of similar biometric records in order to determine the identity of the owner of the biometric record. It is a one-to-many comparison.

Verification is a term used to describe the process of confirming that a person is who he or she claims to be by matching the person's biometric record against that of the claimed identity. It is a one-to-one comparison.

The IQS study is concerned exclusively with the identification function.

2.1.2 Performance Measurements

Measurements of Correct Matches: Reliability and False Reject Rate

Reliability is the probability that a matcher system will correctly identify a search print's mate when the mate is present in the system repository. The complement of reliability is ***False Reject Rate (FRR)***:[3]

$$Reliability = 1 - FRR$$

Thus, a reliability of 99 percent is equivalent to an FRR of 1 percent: both mean (for example) that of 1,000 searches for which matches are present, there will be 990 correct matches and 10 false rejections.

Reliability has substantial operational implications that differ based on the type of system being used:

- For systems in which it is in the subject's interest to be matched (such as a system identifying valid visa holders), falsely rejected individuals can be expected to complain, thus requiring a separate resolution process. In such systems, the operational reliability rate can be determined.
- For systems in which it is *not* in the subject's interest to be matched (such as a criminal history system), falsely rejected individuals cannot be expected to complain: the *operational* reliability rate is generally unknown for such systems, so the tested reliability becomes particularly important.

This report will discuss matcher performance in terms of both FRR and reliability, as appropriate.

[3] FRR is also known as "false negative" or False Non-Match Rate (FNMR).

Measurements of False Matches: False Accept Rate and Selectivity

False Accept Rate (FAR) is the probability that a system will incorrectly determine that a search print and a file print are mates.[4]

The implications of FAR differ for verification and identification systems:

- A verification system with a FAR of 1 percent would mean that of 1,000 individuals presenting forged documents for entry into the United States, 990 would be rejected and 10 would be allowed entry.
- An identification system with a FAR of 1 percent would mean that for *each* individual searched against a database of 40 million subjects, an average of 40,000 false matches would be identified.

Acceptable FAR levels are very different for verification and identification. For that reason, most AFISs also note the "operational FAR," which is known as ***selectivity*** in FBI terminology. Selectivity is the number of false candidates, on average, that would be returned for every search. Selectivity is calculated by multiplying the FAR by the size of the database:

$$Selectivity = FAR * DatabaseSize$$

In general, policy decisions determine acceptable selectivity levels, and the acceptable FAR levels are calculated from these. For example, if it is determined that only 1 search per 100 can return a false match against a repository of 40 million subjects (the size of the FBI's CMF), then the maximum acceptable selectivity would be 0.01, corresponding to a FAR of 2.5×10^{-10}.

Since FAR must be very low for identification systems, even using large numbers of test cases, complex statistical projection methods are required to determine FAR. For example, if 2,000 subjects in a test set are searched against a database of 70,000, the minimum measurable FAR is $7.1 \times 10^{-9} \pm 1.4 \times 10^{-8}$ (95 percent confidence interval).

FAR has substantial operational implications that differ based on the type of system being used:

- For systems in which it is in the subject's interest to be matched (such as a system identifying valid visa holders), falsely matched individuals may be unlikely to complain: the operational FAR or selectivity is generally unknown for such systems, so the projected FAR, based on analysis of test data, becomes particularly important.
- For systems in which it is *not* in the subject's interest to be matched (such as a criminal history system), falsely matched individuals are likely to protest and therefore a separate resolution process must be necessary.

This report will discuss matcher performance in terms of both FAR and selectivity.

[4] FAR is also known as "false positive" or False Match Rate (FMR).

Measurements of Unusable Data: Failure to Enroll

Failure To Enroll (FTE) refers to a fingerprint image of such low quality that it cannot be matched. FTE is closely related to FRR and may be considered a subset of FRR: the false rejects are the cases that do not match, while the FTEs are the cases for which matching is impossible, very unlikely, or impractical.[5] The distinction is that FTE cases can be identified at the time of capture, and the subjects can be directed to some secondary processing system.

Image Quality Metrics (IQMs) are used to determine whether a particular fingerprint is FTE. IAFIS uses equivalent number of minutiae (an image quality measure discussed later in this report) to determine FTE: images with nine or fewer equivalent minutiae are rejected. Approximately 2.5 percent of IAFIS searches are considered FTE.

When analyzing performance, it is important to note whether FTE is included in FRR. In some studies, FTE is reported separate from FRR, and the good—but misleading—FRR is occasionally quoted out of context.

FTE can be caused by natively poor fingerprints (such as from abraded or scarred fingers), poor fingerprint images (such as caused by malfunctioning scanners or poor operational procedure), or by a variety of procedural errors. FTE is likely to increase with some populations, such as women, children, and the elderly. FTE can be expected to increase with poorly trained operators, poorly maintained equipment, or frantic working conditions. Policies can and should be implemented to minimize FTE, but it is naïve to expect that FTE can be reduced to zero in any large-scale operational system.

FTE has clear operational implications: alternative processing must be available for FTE cases. Traditionally, demographic information (name, color of eyes, date of birth, etc.) and identification cards have been the only methods of resolving FTE. Recently, there has been increasing discussion of using non-fingerprint biometrics—such as facial or iris recognition—to resolve FTE cases.

Lights-out vs. Manually Verified Identification

All fingerprint matching is probabilistic. As discussed above, FRR and FAR are the probabilities that an identification system will make an error, either by failing to make a match when one exists or by making a match when one does not exist. In addition, FRR and FAR are not independently adjustable variables; improving one will worsen the other.

A fingerprint identification system can operate in either "lights-out" mode or as a manually verified system. In a lights-out system, a threshold will have to be established for matches. Establishing that threshold determines the trade-off between FRR and FAR for that system.

IAFIS is a manually verified system. IAFIS has two thresholds: one for definite matches and a lower one for questionable matches. Questionable matches are referred to human fingerprint examiners for verification. These thresholds were established after extensive operational experience.

---

[5] Very poor fingerprints often take a disproportionate share of system resources even when they return no matches. For systems in which throughput is critical, very poor images may be considered FTE if they are impractical but not impossible to search.

A manually verified system can have better tradeoffs between FRR and FAR than a lights-out system. The human verification acts as another stage of matching that further separates matches and non-matches.

However, operational error rates cannot be ignored. Humans do make mistakes. Confirming decisions of an automated matching system is a boring and repetitive task. Decisions made by poorly trained, bored, or overworked staff are likely to be *worse* than automatic AFIS decisions.

Using non-fingerprint biometrics—such as facial or iris recognition—is an alternative approach to handling indeterminate matches, which could improve the tradeoffs between FRR and FAR in a lights-out system.

Trade-offs: FRR, FAR, FTE, Responsiveness and Staffing

Discussions of fingerprint matching system performance usually focus on the so-called accuracy of the system: how many true or false identifications are actually made. The reason for this focus is clear. The system designer wants to know how many potentially dangerous persons are not identified or how many persons who are innocent are designated as possible criminals or dangerous aliens.

However, this concern is only a part of the problem. Specifying required values for FRR, FTE, and FAR will drive the cost of the system hardware and will strongly impact system responsiveness and the required level of operational staffing. Trade-offs exist between error rates, staffing and system responsiveness:

- Lowering FRR allows fewer potentially dangerous persons to escape identification, but raises FAR, increasing the need for manual processing of exceptions.
- Manual verification reduces both FRR and FAR but increases staffing requirements and degrades system responsiveness, especially at peak times.
- Reducing FTE increases exception handling requirements, increasing system cost for automated handling, and staffing requirements for manual processing.

It is important that the performance requirements that are specified include consideration of staffing impact.

FTE has an operational staffing impact for most systems. Fingerprints typically exhibit an FTE of about 2 to 5 percent. Assuming an FTE of 3 percent, a daily workload of 20,000 registrations, and 10 minutes per subject for secondary processing, then 100 hours per day of additional staff time will be required to handle FTE cases. In a case like this, the effective FTE could be lowered somewhat by increasing system cost, for example by using better scanners or by using better algorithms such as the IAFIS latent matcher for the worst quality images.

For systems such as IAFIS that use human verification of matches, FAR has a substantial impact on staffing. Assuming a FAR of 0.000002 ($2x10^{-6}$), a database size of 40 million, 65,000 searches per day, and 3 minutes of manual processing per false match, it would take *260,000 hours per day* for manual processing. Reducing the FAR to 0.00000000003 (3×10^{-11}) results in a staffing requirement of *3.9 hours per day*. This is the current performance level of IAFIS for ten-finger-rolled searches of the CMF.

The decision of which operating points to choose are policy-level decisions that must be made by the decision makers that are responsible for implementing the system. They will have to trade off cost—both hardware and staffing—with system performance.

2.1.3 Types of Fingerprints: Rolled, Flat, Slaps, and Latent

Four types of fingerprint images are discussed in this report: rolled, flat, slaps, and latent. Each type of fingerprint image is acquired in a different way and, as a result, has different implications for fingerprint matching and AFIS performance. Each of these four types is discussed in greater detail in the remainder of this section.

Rolled Fingerprints

Ten rolled fingerprints are used by the FBI's IAFIS system for background checks. Rolled fingerprints are generally between one and two and a half square inches and contain an average of 80 minutiae. In general, rolled fingerprints have sufficient ridge detail to allow classification in almost all cases. Figure 1 shows an example of a rolled fingerprint.

Rolled fingerprints provide a great deal of information allowing for highly accurate searches. However, capturing a properly rolled fingerprint is a slow process that requires trained staff, and the operator's manipulation of the subject's fingers often makes the subjects feel "manhandled."



Figure 1. Sample Rolled Fingerprint

Flat Fingerprints

Flat fingerprints—sometimes referred to as "plain" fingerprints—can be captured quickly using inexpensive scanners by individuals with minimal training. Flat fingerprints are generally about 0.5 square inches and contain, on average, 40 minutiae. Figure 2 shows an example of a flat fingerprint.

Flat fingerprints are more difficult to classify than rolled fingerprints; they can be classified less than half of the time using the IAFIS classification algorithms. Flat fingerprints are often associated with inexpensive capture devices, which are typically not of the quality required by AFIS; as a result, there are additional quality implications.



Figure 2. Sample Flat Fingerprint

Slap Fingerprints

Slap fingerprints, or "simultaneous plain impressions," are simply multiple flat fingerprints captured at the same time. Like flat fingerprints, slap fingerprints have an average of about 40 minutiae per finger and can be fully classified less than half of the time. Figure 3 shows an example of slap fingerprints.

The rapid capture process for slaps is as straightforward as for flats, and requires little training. Capturing multiple fingerprints simultaneously is much less prone to error than separately capturing individual flat images: capturing eight fingerprints requires two slap impressions, but eight flat impressions.

Currently, only higher-quality scanners have the large platens required for slaps. As a result, slap images tend to have better image quality than flats.

Slap fingerprints can be acquired in two images (including four fingers in each image and ignoring the thumbs) or three images (including the two thumbs, side by side in the third image). As will be discussed later in the report, the use of multiple fingers significantly improves AFIS reliability and dramatically reduces the cost of hardware.

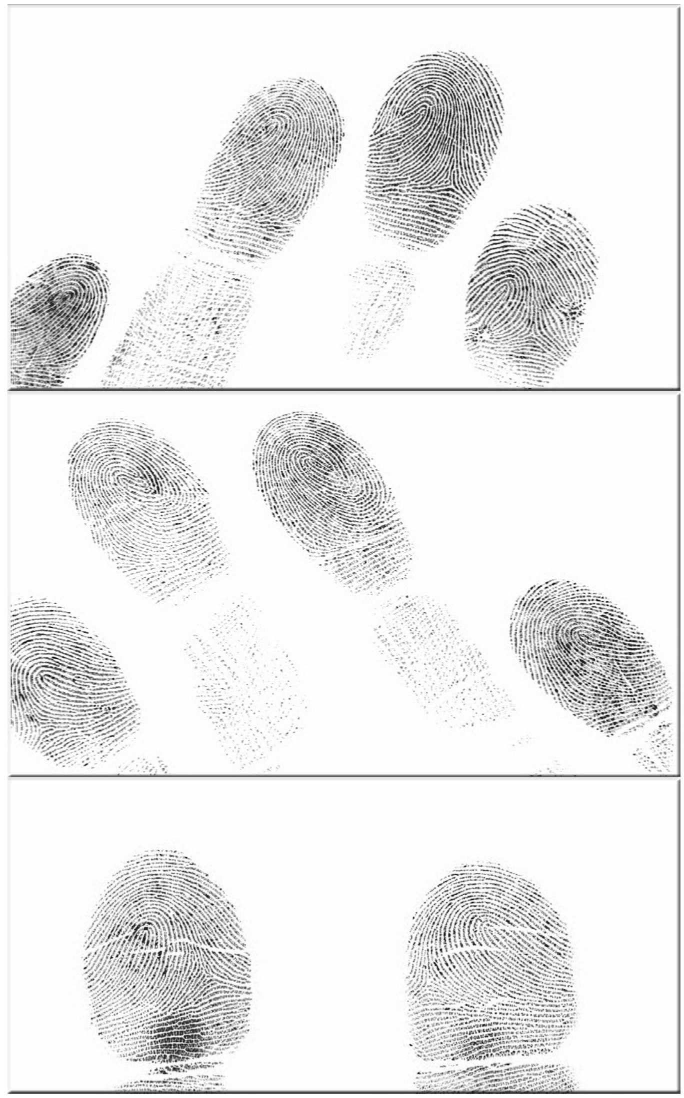


Figure 3. Sample Slap Fingerprints

While slap fingerprints are a reasonable compromise between rolled and flat fingerprints, they are not a panacea. A number of issues must be addressed in order to use slap fingerprints in an operational system:

- ***Image size:*** Standard slap images on paper fingerprint cards are 3" wide by 2" high. Parts of the fingerprints are frequently cropped at that size. Systems that rely on slap fingerprints for identification should set standards requiring larger scanner platens.
- ***Segmentation:*** Slap fingerprints must be segmented to extract the individual fingerprint images from the single large image. Segmentation can introduce errors, increasing FTE or FRR. See Section 8.1 for a more detailed discussion of this topic.
- ***Database Searching:*** Many AFIS systems are tuned to be most effective when searching rolled fingerprints against rolled fingerprints. Systems will have to be reengineered or tuned to maximize accuracy when searching slaps.

Latent Fingerprints

Latents are fingerprints that are unintentionally left on surfaces such as paper and walls as a result of normal handling. Figure 4 shows an example of a latent fingerprint.

Latent searching and identification requires great expertise and is very computer-intensive. Searching 700 latents on IAFIS requires roughly the same resources as searching 40,000 rolled ten-prints. Latent algorithms could be used to search flat or slap fingerprints. As discussed in Section 4.4, the universal application of the IAFIS latent algorithm set is probably cost-prohibitive without reengineering the process within AFIS. However, limited use of the latent algorithm set for poor-quality fingerprints could be a practical alternative. Nevertheless, lessons learned about processing poor-quality latent fingerprints can be applied in new AFIS implementations for processing poor-quality flat or slap images.

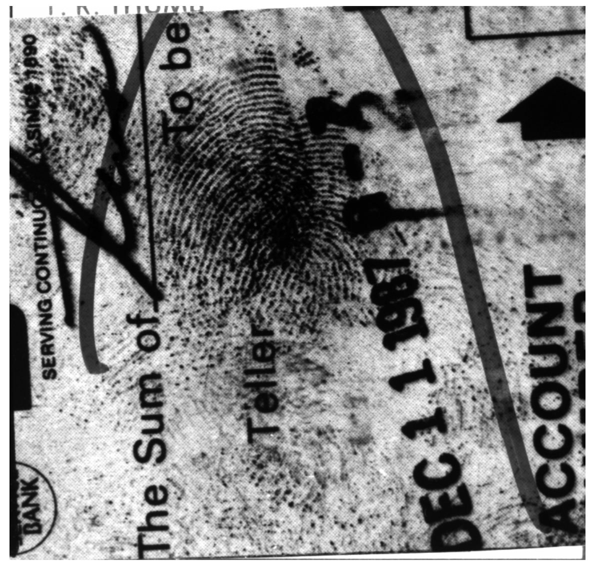


Figure 4. Sample Latent

2.1.4 Image Quality

The term *image quality* can be defined in many ways depending upon the context in which it is used. In a generic technical sense, the quality of an image is defined in terms of such parameters as resolution, contrast, and distortion. These terms are used to describe how faithfully the image depicts the original subject matter.

For the purposes of the IQS, *image quality* refers to the quality of a fingerprint image. In this case, quality is synonymous with image information content (distinguishable patterns and features) that is useful for matching a search print with a file print. The quality of livescan fingerprint images is a complex function of many factors, including the livescan image capture capabilities, the scanning environment and the ambient humidity, the pressure with which the finger is applied to the platen, the cleanliness of the platen, etc.), the state of the fingerprint (severity of scarring, recent abrasion, finger cleanliness, etc.), and the fingerprint area captured. Other factors that can limit fingerprint image quality include the age, sex, and occupation of the subject, the size of the fingerprint, and how the fingerprint has been processed (for instance, inappropriate recompression of the image).

2.2 Algorithm Test Bed Overview

As developed by Lockheed Martin, the ATB provided performance measurement and test capabilities. The ATB concept was refined considerably for the IQS in order to provide an integrated and extensive data collection capability and to minimize test time. The ATB generated numerous reports to facilitate analysis. The ATB performance projection model can use a relatively small test database to make accurate predictions of algorithm performance for a 40 million subject operational database. These capabilities made the ATB a useful tool for simulating INS searches of IAFIS and gaining extensive insight into the underlying search processes.

The ATB included a 70K subject repository, which is a subset of the CMF and contains rolled fingerprint images for approximately 70,000 subjects (ten fingerprint images per subject). During the development of the AFIS segment of IAFIS, this data was used extensively for system testing and tuning and was determined to be representative of the CMF.

While the ATB was essentially the same as the operational AFIS, there are several differences. Some of the ATB latent algorithms had been improved as part of the FBI's Technology Refreshment Program (TRP). These improvements were added to the operational IAFIS after the completion of the IQS. Because of ongoing TRP within IAFIS, IAFIS now uses some algorithms that have been improved since IQS; therefore, operational IAFIS performance should be better than was measured on the ATB for IQS.

In addition, the ATB as used in support of the AFIS development required off-line report generation. For the IQS, LMC wrote a number of scripts that allowed direct linkage to the report generation programs.

An additional difference between the ATB and the operational AFIS environment was that the ATB uses an HP N-Class processor, while IAFIS uses older Convex 2000 technology. All of the ATB software was ported to the N-Class as part of the TRP process. It should be emphasized that no new software was developed in support of this effort. All algorithms and search processes were the same as developed in support of the AFIS development and the AFIS TRP.

2.3 Approach

This section discusses the approach for the IQS study and the Slap Segmentation Study. The results of the IQS study approach are discussed in Sections 3 through 6 and Section 8. The results of the Slap Segmentation Study are discussed in Section 7.

2.3.1 IQS Study Approach

This section provides a brief overview of the approach taken to conduct the IQS. Greater detail is provided later in this report.

The most direct means of determining how IAFIS performs when searched with INS data would have been to search a representative sample of INS data against the operational IAFIS and analyze the resulting performance data. However, this approach would have interfered with IAFIS operations. Instead, an alternative approach was used. Representative INS data would be searched against the ATB.

Repository size has a considerable effect on AFIS performance. Statistical techniques must be employed to use the results obtained from searching a relatively small data repository and project the results that would have been expected had a much larger repository been searched. To perform this projection, the ATB included the extreme-value-statistics projection method used for similar purposes during AFIS development.

The ATB provides a flexible means of estimating IAFIS performance when searched with a variety of fingerprint image types. The typical mode of operation is to employ sets of test data that contain both *search* prints and corresponding (mated) *file* prints; these prints are associated with one another by means of a "truth table." Unmated prints may also be searched to provide filter rate and selectivity data. The file prints are searched against the data repository to ensure that mates to the intended file prints do not already exist in the repository, and any unintended mates are noted. The file prints are then *seeded* into the repository. For the purposes of the IQS, the file prints are rolled ten-prints. Search prints, which can originate as rolled or flat, inked or livescan images, are then searched against the seeded repository.

Each mated data set was searched against the seeded data repository, and the unmated data set was searched against the unseeded repository. The search was performed first using algorithms for searching ten-prints. The search was performed again using algorithms for searching latent fingerprints. The results obtained with each algorithm suite were compared. Searches that missed an expected mate or matched an unanticipated mate were reviewed by a fingerprint expert to determine whether the unexpected result was due to an error in the truth table, a matcher error, or some other anomaly.

As discussed in Section 2.4, the government provided a variety of test data sets for use in this study in an attempt to overcome known data biases and produce realistic, consistent results. In addition to providing several sets of INS fingerprints, a set of mated, rolled ten-print images was provided to baseline ten-print performance of the ATB and compare it with the known performance of IAFIS. This data set was also useful for developing insight into performance expectations when combinations of fingerprints other than a full set of ten-prints or two index-finger prints are searched against IAFIS.

Mitretek developed custom scripts to extract and derive the desired image quality metrics and performance data from the ATB data and reports and to organize them to aid further analysis. These analyses are described in subsequent sections of this report.

2.3.2 Slap Segmentation Study Approach

In December 2000, a limited analysis was conducted of the performance of simultaneous plain livescan impressions (commonly known as "slaps") using Aware segmentation software and the ATB. This study was conducted during the last few days that the ATB was available. The scope of the results is therefore limited to tests run during a very limited time period. The results were sufficient to be meaningful, but in several areas, further testing would have been useful.

Segmentation software improved dramatically after the slaps study was completed. In February 2002, the same data set was segmented using a later commercial release of the Aware software, with substantially better accuracy. Since the ATB was not available for

this test, search performance was not measured. An appendix to the original slaps study document was added to report the February 2002 results.

The data set used was the "Civil 382" (see Section 2.4.5). This data set ***does not*** purport to be representative, but it ***is*** a sample of civil IAFIS submissions and does include a number of poor-quality images. It should be noted that the subjects were captured using livescan fingerprint devices: slap images captured by scanning a paper fingerprint card (even if that card had been printed from a livescan source) have different characteristics. While thumbprints were included in the data set, they were not tested due to time constraints.

The December 2000 test was conducted in two parts: the slap images were segmented using a variety of different methods of preprocessing in an effort to improve performance; the resulting individual fingerprints were searched against the ATB. In the February 2002 test, only segmentation performance was tested.

The slap images were segmented and rotated into separate upright fingerprint images using Aware segmentation software. The Aware software was selected solely due to its availability; it only provides a data point for reference and has not been compared with other segmentation software.

- In the December 2000 test, the slap images were processed using Aware's "FingerPrint Image Segmenter" software (Beta v. 12/01/2000). Since segmentation performance was mediocre, multiple tests using a variety of preprocessing were conducted, including rotating the slaps image before segmentation and saturating the image (forcing very light gray parts of the image to white. The preprocessing clearly improved results.
- In the February 2002 test, the same data set used in the earlier tests was segmented using Aware's *Ten-Print Sequencing Library for Windows, v1.24* (released June 21, 2001). This software performed very well without any image preprocessing.

Determining when segmentation failed was a problem. In the worst cases, the Aware software noted that fewer than 4 fingers per image were found, but the other problem cases could only be identified through visual inspection. Although the software returned a segmentation quality result, it did not provide a meaningful way to distinguish successful and unsuccessful segmentations. In the December 2000 test, samples of the segmentation results were visually inspected; in the February 2002 test, every image was visually inspected.

The search procedure used was the same as for the Image Quality Study. The rolled file fingerprints (cropped to 500 x 500 pixels) were seeded into the 70K background database and searched using different combinations of fingers against the ten-print ATB algorithm suite.

2.4 Test Data Sets

It is difficult to develop data sets for testing biometric systems. Data sets must be representative of a population;, at the same time, they must be mated to allow for testing. One particular problem that results is "survivor bias"—the incorrect assumption that the subjects remaining at the end of the process of developing a data set are representative of the subjects that started the process. Mated data sets are necessarily suspect; the process that determines mates almost always biases the results. In particular, data sets mated by an AFIS are only representative of fingerprints that can be matched by an AFIS.

To test the performance of IAFIS when using IDENT search data, a sample of mated IDENT search data was required. Three samples of INS data and two sets of FBI ten-print data were examined as part of the IQS study:

- Data Set 1 (DS1) and Data Set 3 (DS3) contained flat fingerprints and rolled mates, obtained from the INS.
- Data Set2 (DS2) contained unmated flat fingerprints, obtained from the INS.
- BDM 3520 contained rolled fingerprints used for baseline testing, obtained from the FBI.
- Civil 382 contained rolled and slap fingerprints from civil (applicant) livescan submissions, obtained from the FBI.

Each of these data sets is described in greater detail below.

2.4.1 Data Set 1 (DS1)

DS1 is an INS test data set previously assembled for benchmarking IDENT matchers. It consists of 1,678 index-finger pairs of flat livescan images together with their rolled mated pairs. The test data was intended for determining the image-quality characteristics of both the search and file data and also for determining the reliability and selectivity of IAFIS when searched with IDENT data.

The INS collected this data in the mid-1990s. The collection process included collecting rolled fingerprints from a sample of subjects at the same time the flat fingerprints were collected. This method of collecting mated data was excellent because it was not biased by applying a matcher in the mate-selection process. However, some fingerprints were removed from the original data set. The *PMA-3 Benchmark Test Report* states that two fingerprint experts winnowed the data, removing 16 percent of the subjects; this means that mated, poor-quality fingerprints were not available for testing. The IQS Team removed one subject because one of the flat images was blank and the subject was not noted as an amputee.

2.4.2 Data Set 2 (DS2)

DS2 is a set of ***unmated*** fingerprint pairs that is representative of the operational IDENT input stream. The key requirement and purpose for DS2 was that it be representative of paired index-finger prints obtained at INS border patrol stations and ports of entry. To achieve this goal, the fingerprints were selected so that the data set would exhibit the same image-quality distribution that characterizes the INS's Recidivist file. This data set was termed Data Set 2 (DS2). It contains 2,589 unmated flat index finger pairs. Since DS2 is a representative data sample, its image quality profile is considered a baseline. IQS adjusted

DS1 performance measurements to correct for DS1 variations from the DS2 image-quality profile, as will be discussed in Section 5: Impact of Image Quality on Performance.

2.4.3 Data Set 3 (DS3)

To further ensure that the study employed representative test data, INS compiled a third data set composed of mated pairs. This data set was called Data Set 3 (DS3). DS3 consists of 2,005 mated flat index-finger pairs. DS3 was collected by using the IDENT matcher to determine the mates. Furthermore, of the mated pairs identified, generally those pairs with the highest matcher scores were selected. This resulted in a data set that was dramatically skewed toward matchability since the usual distribution of hard-to-match mates was not included in the data set. This bias rendered the data set virtually unusable.

2.4.4 BDM 3520

In addition to the IDENT data, a data set of rolled ten-prints was also chosen for testing. This test set is known as the Basic Demonstration Model (BDM) data set (BDM 3520). The BDM 3520 contains rolled ten-print images for 3,520 subjects, of which 1,825 subjects have mates in the 70K Repository. It was used extensively for testing during the IAFIS development effort. The characteristics of the BDM data set are well known and understood. It is therefore a good test set for verifying ATB processing behavior and for baselining ATB operations and ensuring the ATB mirrors IAFIS when searched with rolled ten-prints. The data set was also found to be useful for determining the performance of various rolled multi-finger search combinations.

2.4.5 Civil 382

The FBI provided 382 Electronic Fingerprint Transmission Specification (EFTS) files from a sample of civil livescan submissions to IAFIS in January and February 2000; each contained ten rolled fingerprints, two four-finger slaps, and two flat thumbprints. The rolled fingerprints were used to seed the ATB database (against a background database of approximately 70,000 subjects). This data set does not purport to be representative, but it is a sample of IAFIS submissions and does include a number of poor-quality images. It should be noted that the subjects were captured using livescan fingerprint devices; slap images captured by scanning a paper fingerprint card (even if that card had been printed from a livescan source) have different characteristics.

Section 3: Image Quality Metrics

Image Quality Metrics (IQMs) are used to predict the behavior of the search process in relation to fingerprint quality. IQMs can provide information on how fingerprint capture devices and other operational exigencies can have an impact on performance. A detailed understanding of the fingerprint quality distribution to be used in a fingerprint system is critical when making engineering and tuning decisions.

The relationship between fingerprint quality and matcher performance is clear but imperfect, as discussed in Section 4: Performance Measurements. This section focuses on those results that have the greatest bearing on matcher performance. A more complete discussion on this topic can be found in the IQS.

AFIS matching is based on fingerprint classification, topology, and minutiae. Therefore, fingerprint image quality needs to measure the following:

- Classifiability
- Ridge area, definition, and clarity
- Minutiae number, definition, and clarity

IQMs that predict performance of a specific system must be tuned for that system. Prediction of performance against a hypothetical system requires a broader range of metrics.

The IQMs provided by the ATB were developed by Sagem Morpho and LMC. These metrics were supplemented with additional IQMs developed by Mitretek. All of the metrics were devised to quantify fingerprint-specific image quality and are not concerned with generic image quality, such as noise level, resolution, modulation, and distortion. Some metrics were developed for the specific purpose of image quality, while others are byproducts of other image enhancement software. Image quality metrics are vendor-specific, with the notable exception of number of minutiae. As part of the IAFIS development effort, the IQMs were tested for their ability to predict the performance of many different processes.

Currently, IAFIS uses several of the metrics operationally, including equivalent minutia, compactness, and several classification metrics. These are used to select operating points and as rejection criteria.

The analysis of the IDENT search data shows that the flat IDENT search data is generally of much poorer quality than the rolled data used by IAFIS. Most of this distinction can be attributed to the differences between flat and rolled fingerprints; determining whether the image quality difference is specifically attributable to IDENT is beyond the scope of this study. The number of minutiae that are available to the matcher for searching flat fingerprints is less than half the number available for searching rolled data against IAFIS. Most importantly, flat data is much more difficult to classify as to its fingerprint pattern due to the smaller image area and overall lesser image quality.

Several sample fingerprints from DS2 are included—with all associated IQMs—in Appendix A; they provide good examples of the range of flat fingerprint quality.

The IQMs were measured and analyzed to determine their correlation with one another and their ability to predict reliability. This analysis proved that none of the metrics was a strong predictor of reliability for the flat livescan data; however, they were good predictors for rolled data. Using regression analysis and the correlation measurements between the LMC IQMs, Mitretek developed a Unified IQM (UIQM), which fuses several of the IAFIS IQMs to maximize the correlation between flat fingerprint IQMs and matcher performance.

3.1 Overview of IQMs

Detailed description of all the IQMs considered as part of the IQS can be found in the original IQS report. The LMC, Sagem, and IQS IQMs that were found useful can be divided into several groups:

- Minutiae IQMs
- Contrast IQM
- Ridge Flow IQMs
- Pattern classification IQMs
- Combined quality measures

Table 1 provides a description of the most useful IQMs organized by group.

Table 1. Description of IQMs

IQM Group	IQM	Description
Minutiae	Minutiae	Number of minutiae found by the LMC feature extractor
	Equivalent Number of Minutiae	LMC count of high-quality minutiae that are located near other high-quality minutiae
	Composite (Minutiae quality)	Sagem weighted mean of individual minutiae quality values
	Ellipse	IQS measurement of an ellipse around the area containing all minutiae (in pixels)
	Minutiae Area	IQS measurement of a rectangle around the area containing all minutiae (in square inches)[6]
Contrast	Contrast (Histogram)	Sagem measure of image contrast, or the separability of the image's grayscale values
Ridge Flow	Ridge Flow	Sagem measure of ridge flow consistency
	Compactness	LMC measurement of the 'ellipticality' of the good flow region (penalizing holes or concave areas)
Pattern Classification	Pattern Class Quality	IQS count of pattern class references
	Subclass Quality	Subclass Quality—IQS count of unknown subclass (core/delta) ridge counts
	Subject Reference Count	LMC sum of Pattern Class Quality and Subclass Quality
Combined Quality Measures	Unified IQM TP	IQS measure combining Composite, Ridge Flow, Equivalent Number of Minutiae, and Contrast—provides the best relationship to ten-print algorithm performance
	Unified IQM LT	IQS measure combining Composite, Ridge Flow, Equivalent Number of Minutiae, and Contrast (with different coefficients than Unified Quality TP)—determined by Mitretek to provide the strongest relationship to latent algorithm performance

[6] Ellipse was an IQS measure. Since its units (square pixels) were unintuitive, Minutiae Area has been added; it is simply a rectangle around the minutiae in the fingerprint, measured in square inches.

3.2 IQM Analysis

An analysis of the distribution of IQMs across the data sets employed in the IQS serves two main purposes.

- It establishes an image quality baseline against which to compare the image quality of future test and operational data
- It establishes a means to express and understand variations between study data sets

Several general observations should be noted before reviewing the analysis:

- Flat images are substantially worse than rolled images in almost all IQMs. Most IQMs measure characteristics that are associated with the increased image size or number of minutiae found in rolled data.
- Mated subjects (DS1) are somewhat better than non-mated subjects (DS2) in all IQMs. This is to be expected since the process of identifying mates for a search subject typically excludes a number of poor-quality subjects; in the collection of DS1, 16 percent of the subjects were removed.
- Despite their differences, DS1 and DS2 have generally similar IQM distributions. This is significant because if DS2 were substantially worse than the mated data set, there would not have been a basis for making any meaningful estimates of performance based on image quality.

This section provides an overview of the IQM analysis for the most significant IQMs: Pattern Class Reference and Equivalent Number of Minutiae. A complete analysis of IQMs can be found in Appendix A of the original IQS report.

3.2.1 Unclassifiable Fingerprints

Unclassifiable fingerprints are those fingerprints for which the ATB classifier can make no determination. There are only four pattern classes (arch, left loop, right loop, and whorl), so the Pattern Class Reference Count is set to 4 if the classifier cannot make any determination as to pattern class. This is known as being "fully referenced." This metric shows greater differences between DS1 and DS2 than any other IQM. This is clearly shown in Figure 5.

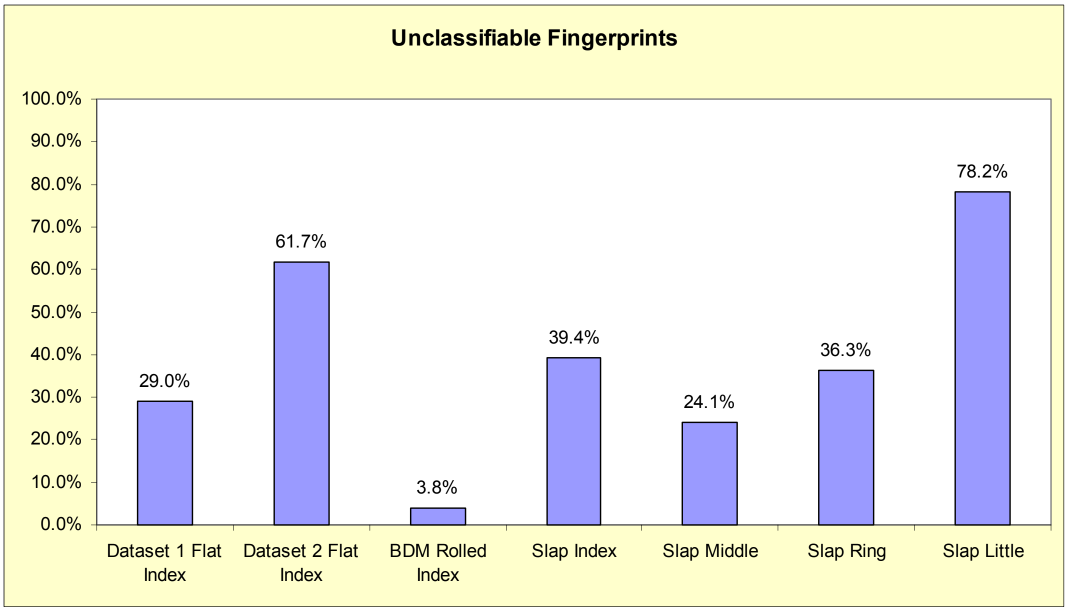


Figure 5. Unclassifiable Fingerprints

(For details, see Appendix B, Table 21.)

A few observations are appropriate:

- The IAFIS classifier is not very effective when used on flat or slap fingerprints. Only about 4 percent of BDM data is fully referenced, far less than any of the flat or slap results. This situation is directly related to the difference in image sizes between flat and rolled images. Flat images often do not include deltas or enough of the ridge structure to clearly determine pattern classification.
- The slap results are from a smaller sample than the flats and rolls, and these results were based on segmentation methods that have subsequently improved, so the slap results are preliminary.
- When slap images are collected, the index and little fingers are partially cropped in many cases. This may explain, in part, why the slap index fingers were harder to classify than the middle or ring fingers and why the little fingers were so much worse.
- Note that thumb results were not available for comparison.

3.2.2 Equivalent Number of Minutiae

Equivalent Number of Minutiae is an LMC count of the minutiae that are determined to be of high quality. Most of the IQMs have distributions similar the distribution of Equivalent Number of Minutiae, which is shown in Figure 6. Only the index fingers from the slaps tests are shown in this chart; the slaps index, middle, and ring fingers had very similar distributions.

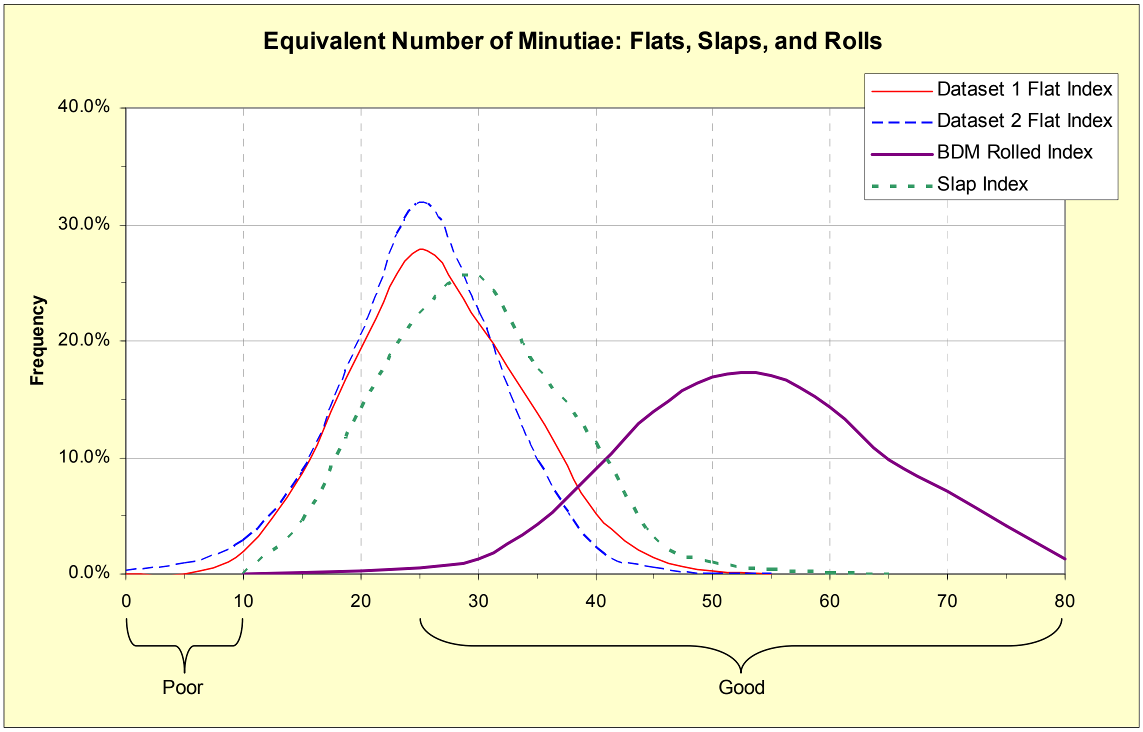


Figure 6. Equivalent Number of Minutiae (Frequency Distribution)

The most salient points to be taken from the Equivalent Number of Minutiae analysis are listed below:

- Rolled images are much better than flat images. Both the number of minutiae and the portion of the minutiae that are considered high-quality are directly related to image size.
- The flat data sets are relatively similar. Although the data sets do show differences, these are outweighed by their similarities.
- DS1 has a slightly greater distribution of good data than DS2.
- DS2 has a slightly greater distribution of poor data than DS1. About 3 percent of the fingerprints in DS2 have fewer than ten Equivalent Minutiae. These fingerprints are very unlikely to match. The current default settings for IAFIS exclude fingerprints in this range as FTE.
- Slap fingerprints are somewhat better than the INS flat fingerprints.

3.2.3 Fingerprint Area

The area of a fingerprint that contains minutiae is represented by Ellipse and Minutiae Area measures. Figure 7 shows the distribution of Minutiae Area for DS1, DS2, BDM, and the slaps index and little fingers.[7] The most obvious characteristic of these distributions is the fact that the dimensions of rolled fingerprints are substantially greater than of flats and that slap index fingers are slightly larger than flats.

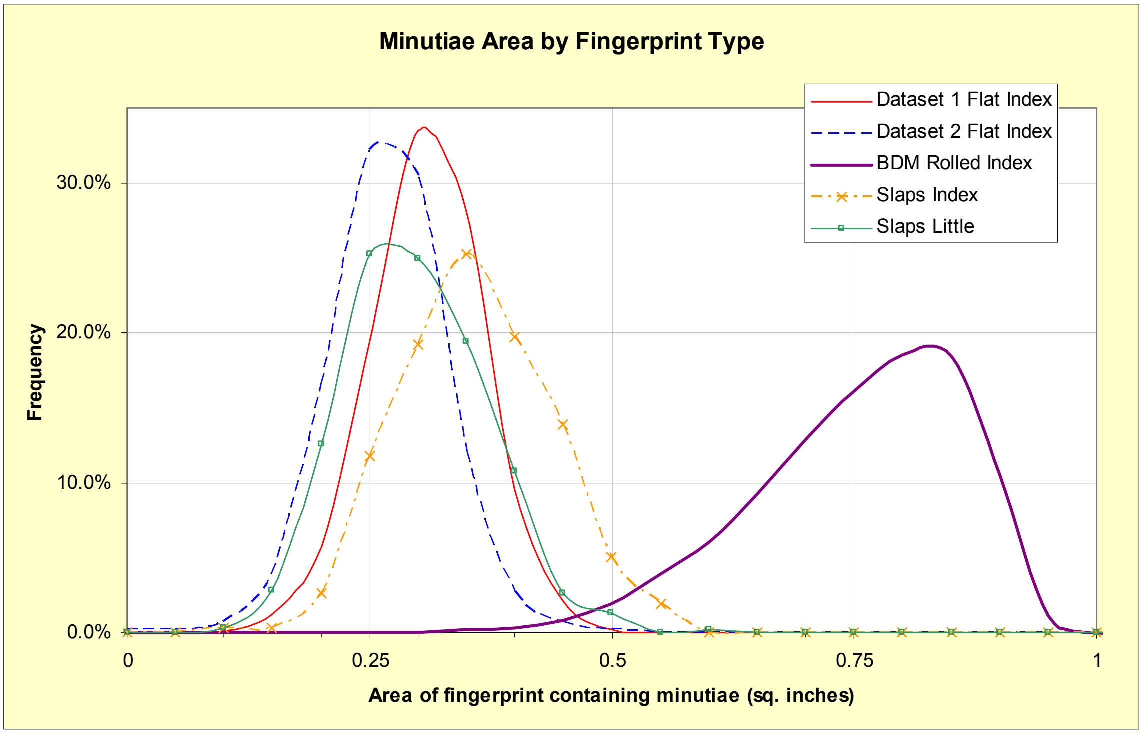


Figure 7. Minutiae Area by Fingerprint Type

3.2.4 Unified Index Quality Metrics

In addition to the measurement of all of the IQMs for each data set, the IQS also analyzed the value of each metric to determine its ability to predict the probability of a successful match. In this analysis, it became clear that a new metric that aggregated several key IQMs was key to quantifying the relationship between image quality and performance. Through a regression analysis process described in Appendix D of the original IQS report, it was determined that the strongest relationship to performance combined the Equivalent Number of Minutiae, Composite, Ridge Flow, and Contrast metrics. For latent and ten-print performance, the same four IQMs are used with different coefficients. The ten-print and latent measures are referred to as UIQM TP and UIQM LT respectively.

Figures 8 and 9 show the distribution of the UIQM (Ten-Print formula). Figure 10 shows detail of the particularly poor quality data.

[7] The distribution of the slaps index, middle, and ring fingers were almost identical; the middle and ring fingers were excluded from the graph for clarity.

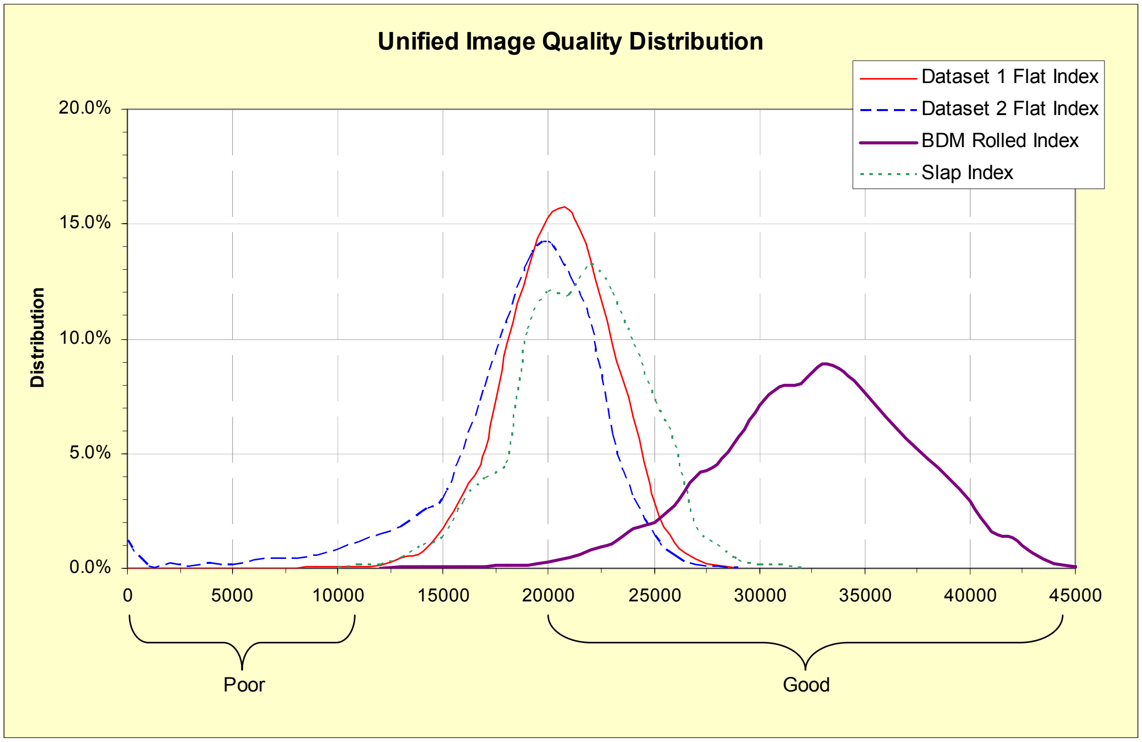


Figure 8. Distribution of Unified Image Quality (Ten-print formula)

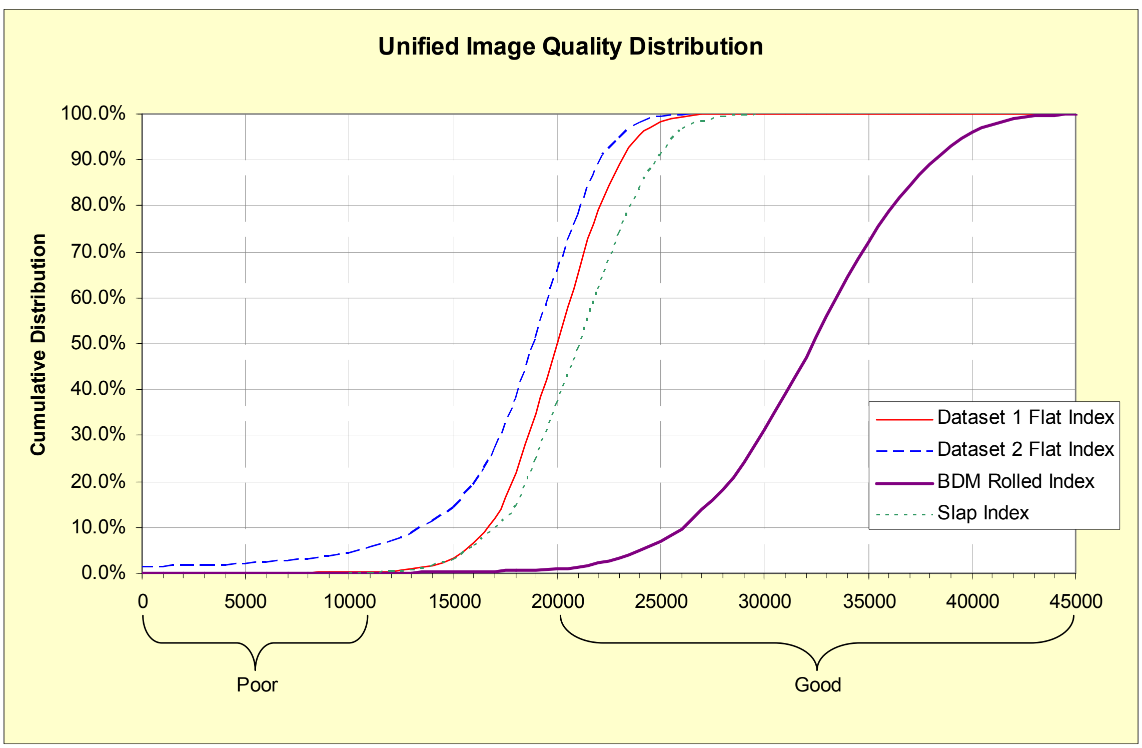


Figure 9. Cumulative Distribution of Unified Image Quality (Ten-print formula)

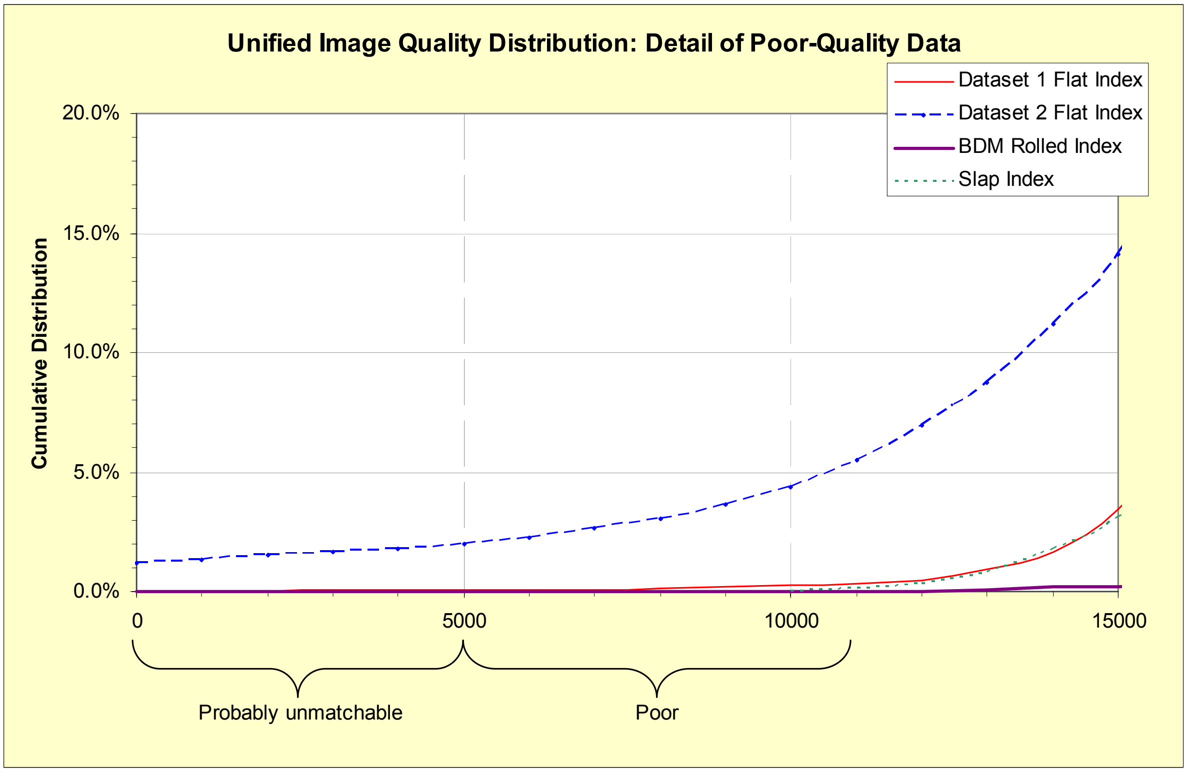


Figure 10. Cumulative Distribution of Unified Quality TP (Detail of Poor Data)

Some important conclusions can be drawn from the analysis of the UIQMs, and it is possible to quantify some previous observations:

- The best flat fingerprints have the same quality as the worst 10 percent of rolled fingerprints.
- About 2 percent of DS2 is so poor that it is virtually unusable (UIQM TP below 5,000).
- Essentially none of DS1 corresponds to the poorest 5 percent of DS2.
- Discarding the worst 5–7 percent of DS2 would leave an image-quality distribution very similar to DS1. (About 16 percent of DS1 was discarded when it was collected.)

The effect of these conclusions will be discussed further in Section 5: Impact of Image Quality on Performance.

3.2.5 Human Review of Poor-Quality Fingerprint Images

The image quality analysis found that images with UIQM TP values below 5,000 were of particularly poor quality. 1.97 percent of DS2 (102 images) was below this threshold. These particularly poor-quality DS2 images were reviewed individually.

Human review of these images revealed the following:

- 0.23 percent of the images (12 images) contained enough fingerprint information that they might be expected to be matched, albeit at a low score. Some of these fingerprints had unusually few minutiae, which perturbed the UIQM calculations.
- 1.26 percent of the images (65 images) were of such marginal quality that it would be extremely unlikely that any automated matcher could make a definitive identification.
- 0.31 percent of the images (16 images) were very poor quality images that would be far beyond the capabilities of automated matchers.
- 0.17 percent of the images (9 images) contained no useful fingerprint information.

This review concluded that 1.74 percent of the images in DS2 should have no expectation of matching in any large-scale identification system.

3.3 Image Quality Analysis Findings

Flat images are substantially worse than rolled images in almost all IQMs. Most IQMs measure characteristics that are associated with the increased image size or number of minutiae found in rolled data. Slap images (except for little fingers) are slightly better than flats in most IQMs.

Pattern classification in particular is likely to be a continuing problem for systems that must process flat and slap fingerprints. The IAFIS classifier is not very effective when used on flat or slap fingerprints. The reliance of IAFIS on pattern classification may have to change; different very fast indexing/filtering mechanisms should be considered that are more effective for flat fingerprints.

Section 4: Performance Measurements

An AFIS is not a single monolithic fingerprint matcher. An AFIS is generally composed of a series of filters and matchers that use fingerprint pattern classification, features, and topology to pare down the huge number of potential candidates, step by step. Generally, fast filtering and rough (prescreen) matchers are used for the initial stages, so that computationally intensive final-stage matchers only have to work on a very limited subset of the database.

Figure 11 shows a simplified overview of the IAFIS matcher, with examples of rolled ten-print and flat two-print searches. The effectiveness of the SSP filter and prescreener is underscored by the fact that for an average rolled ten-print search, the main matcher only has to make a match from 8,000 candidates out of a database of 40 million. When two flat fingerprints are being searched, the matcher workload is roughly 30 times as great as for ten rolled fingerprints.

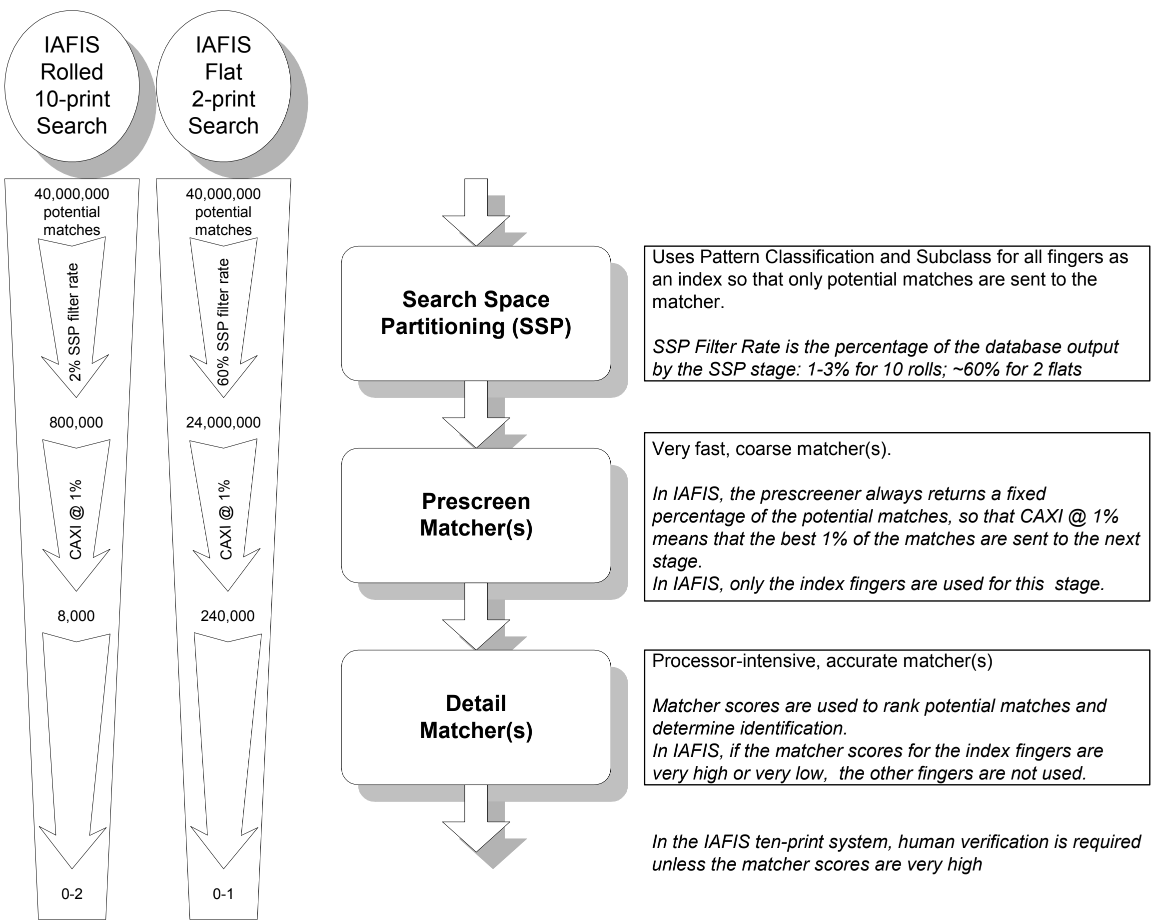


Figure 11: AFIS Matcher Architecture

The IAFIS ten-print system uses a single prescreen matcher and a single detail matcher for high throughput. The latent system does not use search space partitioning; it uses two prescreen matchers in series and two detail matchers in parallel to increase reliability at a great increase in processing requirements.

Even when searching ten-finger data, the IAFIS matcher does not use all ten fingers for every stage of a search. All fingers are used for search space partitioning. Only the index fingers are used for the prescreen matcher. The detail matcher first uses the index fingers for a match; if the matcher score for index fingers alone is very high or very low, a match or non-match decision is made without using the other fingers.

4.1 Search-Space Partitioning Filter Rate

The IAFIS 10-print system classifies fingerprints by pattern classification (left loop, right loop, whorl, or arch) and subclass (ridge count between core[s] and delta[s]). This pattern class is used as an index during the matcher stage known as Search-Space Partitioning (SSP). SSP almost instantly looks up the set of fingerprints in the database that could be potential matches for a specific search. If the complete pattern classification and subclass can be determined for a search, SSP will only send a small portion of the database to the next stage of the matcher; if the search fingerprint cannot be classified, SSP must send the entire database on to the next stage, dramatically increasing its processing requirements.

As discussed in Section 3.2.1, the size of each fingerprint image has an effect on classification, and thereby on SSP filtering. The proportion of unclassifiable fingerprints determines filter rate. If a fingerprint is fully referenced, then the SSP stage of the matcher process cannot filter out any possible candidates. Figure 5 indicates that about 60 percent of the fingers in DS2 will have a filter rate of 100 percent.

More fingers substantially improve the effectiveness of SSP filtering: ten rolled fingerprints (from BDM) have an average SSP filter rate of 1.5 percent, but two flat fingerprints (from DS2) have an average SSP filter rate of over 60 percent. This means that the processing requirements to match two flat DS2-type fingerprints are on average 40 times as great as for ten rolled fingerprints. Table 2 shows the striking improvement that comes from using more fingers, as demonstrated with rolled and slaps data. Note that for each pair of fingers added to the search (other than the little fingers), the filter rate is approximately halved.

The rolled and slaps results both show a dramatic reduction in the filter rates with the availability of additional fingers. This reduction has an important impact on the resource requirements for the search processor. It will reduce the number of special processing cards that would otherwise be required for the prescreen match and also reduce the number of candidates that must be searched by the more computationally intensive final matcher stage.

Table 2: Search Space Partitioning Filter Rates by Finger Combination

	# Subjects	2 Fingers index	4 Fingers index / middle	4 Fingers index / thumb	6 Fingers index / middle / ring	8 Fingers all except thumb	10 Fingers all
Rolled	3520	30.9%	15.0%	10.5%	7.0%	4.4%	1.5%
Slaps	117	46.9%	24.5%	-	13.2%	10.1%	-
Flat (DS2)	2589	61.7%	*30%*	*21%*	*14%*	*9%*	*3%*
			Estimated Values (from IQS)				

(For details, see Appendix B, Table 22 and Table 23)

The ability of a classification algorithm to correctly classify a fingerprint is a fundamental determinant of system requirements. Since the existing IAFIS classification algorithm performs poorly with flat fingerprint data, it will be necessary to improve classification algorithms, and tune them for flat fingerprints.[8]

SSP is generally not a cause for false rejects. Table 3 shows that FRR impact of SSP is minor.

Table 3. Search Space Partitioning Filter Rates and Accuracy

Type	Data Set	Number	Fingers	SSP Filter Rate	SSP Reliability	SSP FRR
Rolled	BDM	2	index	30.9%	99.84%	0.16%
		4	index / middle	15.0%	99.73%	0.27%
		6	index / middle / ring	7.0%	99.62%	0.38%
		8	all except thumb	4.4%	99.40%	0.60%
		10	all	1.5%	99.23%	0.77%
Flat	DS1	2	index	44.6%	99.88%	0.12%
*Note: Reliability/FRR stated here are for the search space partitioning stage, **not** the entire process.*						

4.2 Measured Two-Finger Flat Results Using the Ten-Print Algorithm

The IAFIS ten-print system was designed and tuned to efficiently match sets of ten rolled fingerprints. It combines very high throughput with high accuracy when used with sets of ten rolled fingerprints. When used with two flat fingerprints, the existing ten-print system is not as accurate.

When the ATB ten-print system performs a match, it returns 0 or more possible matches with scores. If the score is above a high threshold, the match is considered certain, with no possibility of a false match.[9] Lower thresholds have increasing odds of false matches. As seen in Table 4, over 97 percent of rolled fingerprint searches have scores above the certain match threshold, but only about 57 percent of two-finger flat searches do. The table also shows that the tuning the ATB settings by changing the prescreen rate increases the

[8] Since the IQS, the pattern classification algorithms used in IAFIS have been improved.

[9] Since IAFIS has now been operational for several years, with human examiners verifying between 40,000 and 80,000 IAFIS matches per day, the distribution of false matches is now very well known. Operationally, IAFIS no longer uses humans to verify very high-scoring matches.

proportion of marginal matches, but has little effect on the number of high-scoring matches.

The table emphasizes the fast drop-off in reliability in terms of matcher scores for the flat livescan data. The drop-off for rolled data is much less severe, which allows greater latitude in selecting operational thresholds rolled data. That trade-off for flat livescan data is not as favorable.

Table 4. Ten-Print System Reliability by Matcher Stage

	DS1 Flat		BDM Rolled	
	Index fingers		Index fingers	All fingers
	CAXI=1%	CAXI=5%	CAXI=1%	CAXI=1%
Passed SSP	99.9%	99.9%	99.8%	99.2%
Passed prescreen	92.1%	96.1%	99.1%	98.8%
Poor Match (Score > 3,200)	89.9%	93.1%	99.0%	98.7%
Marginal Match (Score > 5,000)	87.2%	89.9%	98.9%	98.7%
Good Match (Score > 10,000)	76.3%	77.7%	98.4%	98.4%
Certain Match (Score > 16,000)	57.0%	57.7%	97.2%	97.9%
	1678 subjects		*1825 subjects*	

The accuracy of the prescreen matcher is much worse for flats than for rolls. The prescreener is the largest source of missed identifications. This is detailed in Table 5. False rejects can be reduced by relaxing the prescreener settings, but unfortunately this has a negative effect on the false accept rate, as we will see below. In addition, while raising the CAXI filter rate improves ten-print reliability, it increases the number of matches that must be performed by the final stage matchers. This means the overall system workload would increase.

Table 5. Ten-Print Search Miss Analysis

	DS1			BDM index fingers
	CAXI=1%	CAXI=5%	CAXI=10%	CAXI=1%
Lost in SSP	0.1%	0.1%	0.1%	0.2%
Lost in prescreen	7.9%	3.8%	1.8%	0.8%
Lost in matcher	2.1%	3.0%	*n/a*	0.2%
Poor match score (score < 5000)	4.8%	6.2%	*n/a*	0.0%

The problem with matches with marginal matcher scores is that the proportion of false matches increases, as shown in Figure 12.

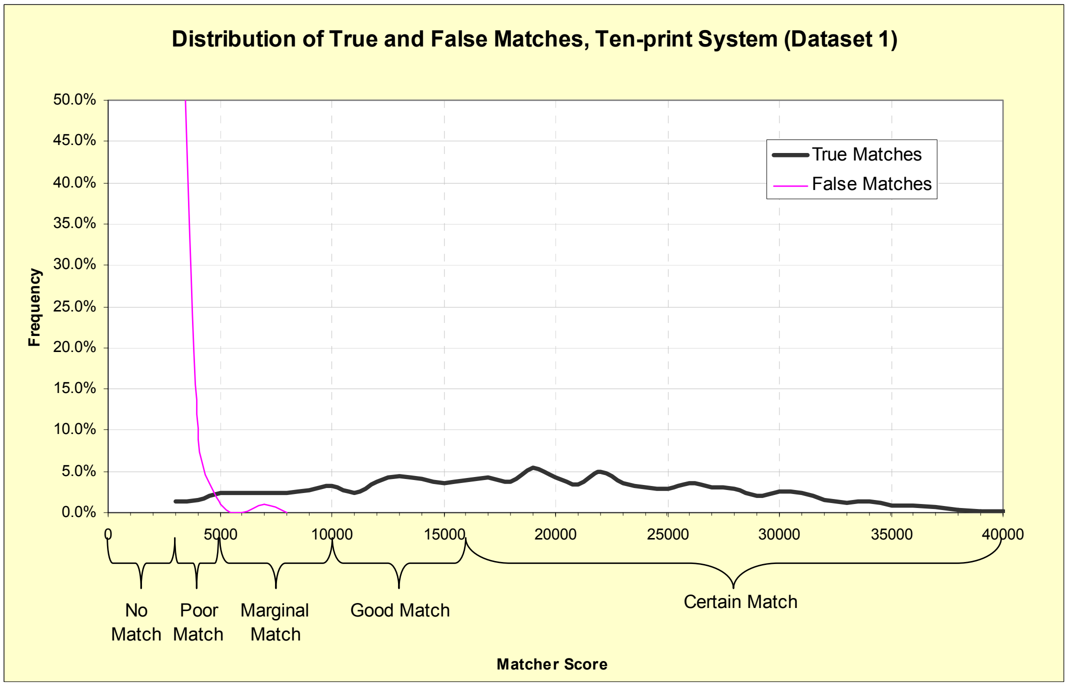


Figure 12. Distribution of True and False Matches in the Ten-Print System

The measured false accept rate is detailed in Table 6. It is important to note that high-scoring false matches are rare, and that therefore with limited test set and database sizes, the presence or absence of a single subject may be cause for a misleading result. In particular, it should be noted that the lack of any false match observations ***does not*** mean that the FAR is zero, but below the minimum measurable level. The minimum measurable FAR and its associated 95 percent confidence interval are noted in each of the charts that report FAR.

Note that the FAR for rolled fingerprints is considerably worse than for flat data. Since rolled images have a greater amount of ridge detail, and an associated greater number of minutiae, areas that have low to moderate degrees of similarity between the search and file print will be statistically more common, so false candidates with low to moderate scores are more common in rolled-rolled searches than in flat-rolled searches.

Table 6. Observed FAR by Matcher Stage (Ten-Print System)

	DS1 Flat Index fingers		DS2 Flat Index Fingers	BDM Rolled	
				Index fingers	All fingers
	CAXI=1%	CAXI=5%	CAXI=1%	CAXI=1%	CAXI=1%
Passed SSP	0.446000000	0.446000000	0.614000000	0.309000000	0.015000000
Passed prescreen	0.004460000	0.004460000	0.006140000	0.003090000	0.000150000
Poor Match (Score > 3,200)	0.000000869	0.000002963	0.000000817	0.000012769	-
Marginal Match (Score > 5,000)	0.000000009	0.000000077	0.000000043	0.000000787	-
Good Match (Score > 10,000)	-	-	-	-	-
Certain Match (Score > 16,000)	-	-	-	-	-
Minimum measurable	*0.000000009*	*0.000000009*	*0.000000005*	*0.000000004*	*0.000000004*
95% confidence interval	*<0.000000026*	*<0.000000026*	*<0.000000014*	*<0.000000012*	*<0.000000012*

Comparing FARs for identification systems is sometimes confusing since differentiating between FARs of 10^{-7} and 10^{-9} is unintuitive. Selectivity is a useful basis for comparison. Table 7 shows the corresponding selectivity for the tests considered above. For example, if the system were used with DS1-type flat fingerprint data at a prescreener setting of 5 percent (the second column), the reliability would be about 89.9 percent and the FAR would be about 7.7×10^{-8}. The selectivity table shows that a FAR of 7.7×10^{-8} would mean that on average three false matches would be returned for every search. Clearly, this would be unacceptable.[10] Determining the worst acceptable selectivity is important. If, for example, it is determined that only one search per hundred could return a false match, then selectivity must be no greater than 0.01. For a database of 40 million, this corresponds to a FAR of 2.5×10^{-10}. In general, FAR of about 10^{-11} is desirable for national identification systems. Projecting performance to this level is discussed in the Section 6.1.

Table 7. Estimated Selectivity against 40M Database by Matcher Stage (Ten-Print System)

	DS1 Flat Index fingers		**DS2 Flat Index Fingers**	**BDM Rolled Index fingers**	**BDM Rolled All fingers**
	CAXI=1%	CAXI=5%	CAXI=1%	CAXI=1%	CAXI=1%
Passed SSP	17,826,620.0	17,826,620.0	24,541,580.0	12,350,730.0	599,550.0
Passed prescreen	178,266.2	178,266.2	245,415.8	123,507.3	5,995.5
Poor Match (Score > 3,200)	34.7	118.4	32.7	510.4	-
Marginal Match (Score > 5,000)	0.3	3.1	1.7	31.5	-
Good Match (Score > 10,000)	-	-	-	-	-
Certain Match (Score > 16,000)	-	-	-	-	-
Minimum measurable	*0.3*	*0.3*	*0.2*	*0.2*	*0.2*
95% confidence interval	*< 1.04*	*< 1.04*	*< 0.56*	*< 0.48*	*< 0.48*

4.3 Measured Multi-Finger Search Results Using the Ten-Print Algorithm

Matcher performance improves when more than just the index fingers are used in matching. Rolled and slap fingerprints were tested in various combinations, as shown in the following tables. It is important to note that the slap results are preliminary: the data sets used were too small to provide definitive results.

These tables show that the relationship between reliability and FAR can be improved as the number of fingers increases. In these tests, the reliability is generally constant: it improves slightly when the middle and ring fingers are added to the index fingers, and degrades slightly when the little fingers are added. As the reliability is kept constant, the FAR drops rapidly beyond the point where it is measurable. These tests are not sufficient to model the effect that adding fingers has on FAR, but it is reasonable to conclude that going from 2 to 4 fingers improves FAR by at least an order of magnitude. Further tests are needed, but these results are encouraging, and provide some indication that the existing ten-print algorithm suite would be able to provide adequate performance using slaps data.

Details for all table are found in Appendix B.

[10] A system such as IAFIS uses human experts to verify matches and prevent the false matches from being reported to the end user.

Table 8. Observed Reliability by Finger Combination for Flat and Rolled Fingerprints

	Flat DS1		BDM Rolled				
	2 fingers	1 finger	All fingers	8 fingers	6 fingers	4 fingers	2 fingers
Poor Match (Score > 3,200)	93.2%	73.5%	98.7%	98.7%	98.9%	99.0%	99.0%
Marginal Match (Score > 5,000)	89.9%	72.1%	98.7%	98.7%	98.9%	99.0%	98.9%
Good Match (Score > 10,000)	77.7%	63.9%	98.4%	98.4%	98.5%	98.9%	98.4%
Certain Match (Score > 16,000)	57.0%	52.8%	97.9%	97.9%	98.2%	98.2%	97.2%
	1678 subjects	3356 subjects	1825 subjects				

Table 9. Observed Reliability by Finger Combination for Slap Fingerprints

	Flat DS1	Slaps Segmentation Method 1				Slaps Segmentation Method 2	
	2 fingers	8 fingers	6 fingers	4 fingers	2 fingers	8 fingers	4 fingers
Poor Match (Score > 3,200)	93.2%	96.5%	97.4%	97.4%	98.3%	100.0%	100.0%
Marginal Match (Score > 5,000)	89.9%	96.5%	97.4%	97.4%	98.3%	100.0%	100.0%
Good Match (Score > 10,000)	77.7%	94.7%	96.5%	95.7%	93.9%	98.6%	97.2%
Certain Match (Score > 16,000)	57.0%	91.2%	92.2%	93.0%	87.0%	95.4%	95.0%
# Subjects	1678	113	115	115	115	218	
Note: Preliminary results for slaps							

In Table 9, note the difference in the reliability at the Certain Match level for 2-finger flats vs. slaps (57 percent vs. 87 percent). Even though the slaps result is preliminary,[11] this is a clear indication of the performance impact of the operational flat data, which used inexpensive scanners. This difference suggests that operational changes such as the type of scanner used may have a substantial impact on system accuracy.

Table 10. Observed FAR by Finger Combination for Flat and Rolled Fingerprints

	Flat DS1		BDM 3520 Rolled				
	2 fingers	1 finger	All fingers	8 fingers	6 fingers	4 fingers	2 fingers
Poor Match (Score > 3,200)	3.0E-06	6.4E-06	-	-	-	4.8E-07	1.3E-05
Marginal Match (Score > 5,000)	7.7E-08	2.0E-06	-	-	-	-	7.9E-07
Good Match (Score > 10,000)	-	1.7E-08	-	-	-	-	-
Certain Match (Score > 16,000)	-		-	-	-	-	-
Minimum measurable	8.5E-09	4.3E-09	4.1E-09				
	1678 subjects	3356 subjects	3520 subjects				

[11] A 95% confidence interval is 87.0% ± 6.2%. See Table 28.

Table 11. Observed FAR by Finger Combination for Slap Fingerprints

	Flat DS1	Segmentation Method 1				Segmentation Method 2	
	2 fingers	8 fingers	6 fingers	4 fingers	2 fingers	8 fingers	4 fingers
Poor Match (Score > 3,200)	2.96E-06	-	-	-	5.13E-06	-	-
Marginal Match (Score > 5,000)	7.71E-08	-	-	-	2.44E-07	-	-
Good Match (Score > 10,000)	-	-	-	-	-	-	-
Certain Match (Score > 16,000)	-	-	-	-	-	-	-
Minimum measurable	8.51E-09	1.22E-07				6.52E-08	
95% confidence interval	<2.52E-08	< 3.61E-07				< 1.96E-07	
# Subjects	1678	117				218	
Note: Preliminary results for slaps							

4.4 Measured Search Results Using the Latent Algorithm

The latent algorithm suite shows a dramatic improvement in reliability as compared with the ten-print system, as shown in Table 12. This improvement is accompanied by a clear improvement in FAR. While the performance is better than the ten-print system, it is still not to the level in which 95 percent reliability corresponds to immeasurable FAR, which would have been desirable.

Table 12. LT Reliability and FAR as a Function of Matcher Score

		Reliability		FAR		
		DS1		DS1		DS2
		LT-CAXI=10%	LT-CAXI=20%	LT-CAXI=10%	LT-CAXI=20%	LT-CAXI=10%
Poor	> 1000	98.1%	98.4%	4.8E-05	5.9E-05	5.8E-05
Marginal	> 1800	95.6%	96.1%	8.6E-09	3.4E-08	2.9E-08
High	> 2500	89.7%	90.6%	-	-	-
Very High	> 3000	83.8%	85.4%	-	-	-
Minimum measurable				*8.6E-09*	*8.6E-09*	*4.8E-09*
95% confidence interval				*< 2.6E-08*	*< 2.6E-08*	*< 1.4E-08*

Table 13. LT Reliability and Selectivity as a Function of Matcher Score

		Reliability		Selectivity		
		DS1		DS1		DS2
		LT-CAXI=10%	LT-CAXI=20%	LT-CAXI=10%	LT-CAXI=20%	LT-CAXI=10%
Poor	> 1000	98.1%	98.4%	1,907.17	2,375.11	2,333.14
Marginal	> 1800	95.6%	96.1%	0.34	1.38	1.15
High	> 2500	89.7%	90.6%	-	-	-
Very High	> 3000	83.8%	85.4%	-	-	-
Minimum measurable				*0.34*	*0.34*	*0.19*
95% confidence interval				*< 1.02*	*< 1.02*	*< 0.56*

4.5 Comparing the Latent and Ten-Print Algorithm Search Results

The latent algorithm clearly performed better than the ten-print algorithm. Figure 13 compares the relative performance, showing the 95 percent confidence intervals for each. Note how the confidence interval bounds are close together at values above 10^{-7} but start to diverge dramatically below that level. Projecting these results to FAR values of 10^{-10} is discussed in Section 6.1.

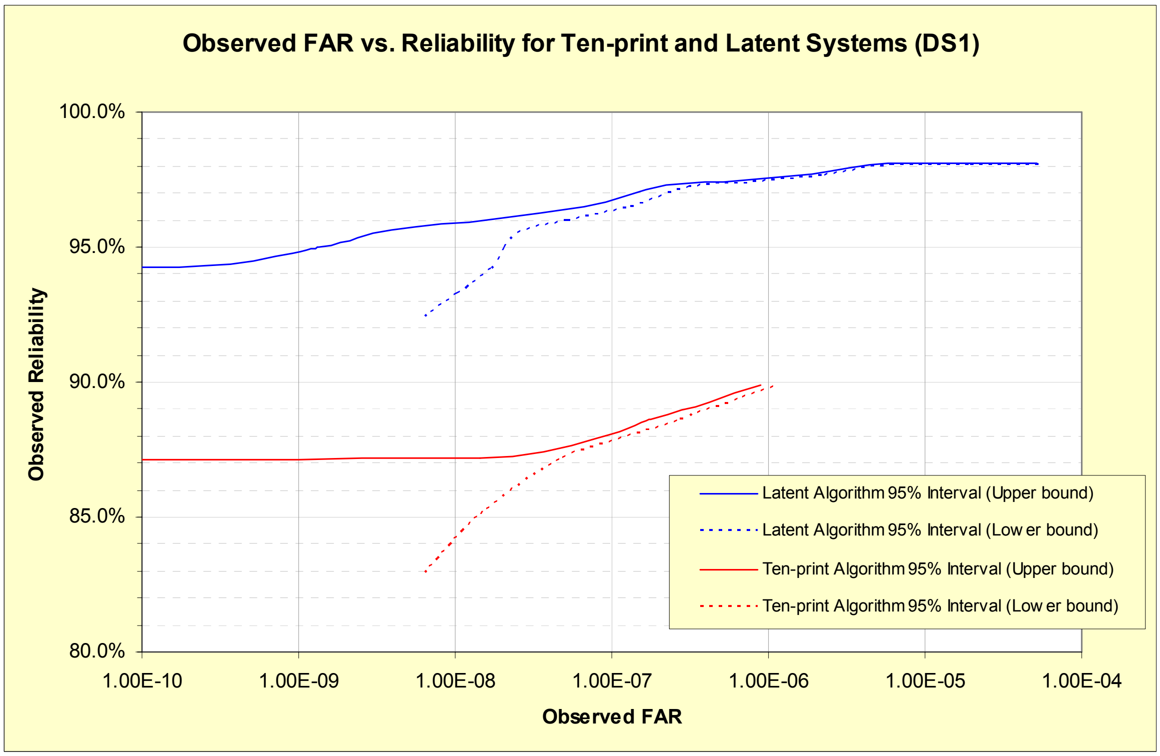


Figure 13. Latent and Ten-print System FAR vs. Reliability for DS1

Note that these results do not account for the difference in image quality between DS1 and DS2, so these results are better than should be expected for operational IDENT data. These adjustments will be explained in Section 5. Note also that these results do not project performance to the 40M database size of the CMF. These projections will be explained in Section 6.

Section 5: Impact of Image Quality on Performance

As discussed in Section 3, the fingerprint quality of DS1 is better than that of DS2. The performance measures reported in Section 4 measured the performance of DS1; the performance of DS2 cannot be measured since it is unmated. This section summarizes the relationship between fingerprint quality and performance, as well as the methods used to estimate DS2 performance.

The relationship between search fingerprint quality and performance is imperfect. The ***best*** of the ATB IQMs (Equivalent Minutiae) only has a 0.33 correlation to the final matcher score. This is shown graphically in Figure 14.

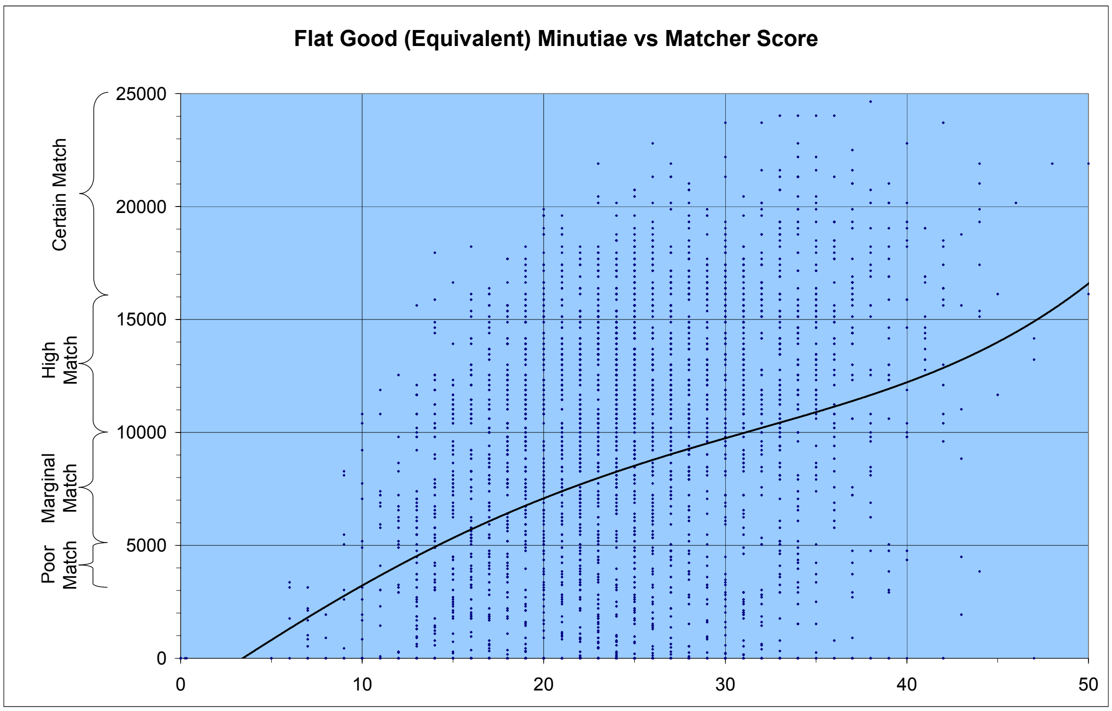


Figure 14. Equivalent Minutiae vs. Matcher Score

The ability to match fingerprints is dependent on three characteristics:

- Number of Fingers
- Correspondence between Search and File images:
    - Overlapping areas
    - Lack of mutual distortion
- Quality of ***both*** Search and File images:
    - Quality of ridge detail
    - Number of features
    - Size of image

Correspondence between fingerprints is a function of the degree of overlap and distortion between the search print and file print, as well as the inherent minutiae content. With a mated set of fingerprints, image quality metrics can be used to quantify the quality of the

search and file prints separately. However, the similarity of the file and search prints is what determines the performance of the matcher; for the FBI's AFIS, it is quantified using matcher scores.

With mated data sets, the quality of the search and file prints can both be quantified; with unmated data sets, only the quality of the search print can be quantified. It is important to note that the image quality of the search image is only one variable in determining how a particular image will perform in an AFIS search. By analyzing the relationship between search and file print qualities and matcher performance, it is clear that the quality of the rolled file print has a greater relationship to performance than the quality of the flat search print. This can be verified by visual inspection of mated pairs of fingerprints, as shown in the following two pairs of images from DS1. The poor quality search image in Figure 15 matched successfully, while the good quality search image in Figure 16 failed to match. Visual inspection clearly shows that in this case the quality of the search print is secondary to the quality of the file print in determining matcher performance.



Figure 15. Poor Quality Search Print That Matched Successfully

Even though the quality of the search print is not the primary determinant of performance for individual searches, in aggregate, search print quality is correlated to performance. In particular, poor search print quality is strongly correlated to poor matcher performance.

Matcher performance, as discussed in Section 4: Performance Measurements, is measured in terms of filter rate, selectivity, and reliability. Although mated data sets, such as DS1, can be used to provide all three performance measures, operational data is only represented in IQS by the unmated DS2. Filter rate and selectivity could be measured directly using DS2, but useful reliability measurements could only be obtained using DS1. By determining the relationships between image quality and matcher score found in DS1 and

adjusting those rates based on the image quality distribution in DS2, a valid estimate of reliability that takes fingerprint quality into account can be derived. This analysis is described in detail in the IQS.



Figure 16. Good Quality Search Print That Failed to Match

The estimation method used to predict the performance levels for DS2 uses a combination of image quality metrics developed by Mitretek called the Unified IQM (UIQM). UIQM fuses several Sagem and LMC IQMs to provide a good predictor of reliability for the flat livescan fingerprint search by IAFIS. These relationships can be applied to DS2 to estimate what would have been the true system reliability. This estimation method was applied to both the ten-print and latent algorithm search results. The predictions for the two algorithm suites are shown in Table 14 and Table 15.

Table 14. Image Quality Adjusted Reliability (Ten-print Algorithm)

	Matcher Score Over …					
	2,000	3,200	5,000	7,000	10,000	16,000
DS1 Measured (CAXI 1%)	90.6%	89.9%	87.2%	83.0%	76.3%	57.0%
Image Quality Adjusted Estimate (CAXI 1%)	86.7%	86.0%	82.4%	75.5%	65.4%	45.5%
DS1 Measured (CAXI 5%)	94.9%	93.1%	89.9%	85.2%	77.7%	57.7%
Image Quality Adjusted Estimate (CAXI 5%)	92.0%	88.8%	84.7%	77.3%	66.6%	46.1%
			Likely operating range			

Table 14 shows that for the ten-print algorithm there is an approximate drop of 5 percent to 11 percent in reliability between the measured DS1 reliability and the image quality adjusted reliability. This difference is in agreement with observations made by INS regarding the amount of data that cannot be matched by IDENT.

Table 15. Image Quality Adjusted Reliability (Latent Algorithm)

	Latent Matcher Score Over …				
	1,400	1,800	2,100	2,400	2,700
DS1 Measured (CAXI 10%)	97.5%	95.6%	93.3%	90.7%	87.8%
Image Quality Adjusted Estimate	93.5%	90.7%	87.3%	83.4%	80.5%
		Likely operating range			

The estimates for the latent algorithm suite (see Table 15) indicate an approximate drop of 4 percent to 7 percent between the measured DS1 reliability and the image quality adjusted reliability. This suggests that neither algorithm can process the very worst data.

Note that these image quality adjustments are very specific to the relationship between DS1 and DS2.

Section 6: Performance Projections

6.1 Performance Prediction for Full-Sized Databases

The difficulty in predicting performance for large databases lies in the fact that an acceptable FAR is orders of magnitude smaller than can be measured during testing. The AFIS repository contains fingerprint information for nearly 40 million people. This is more than 570 times larger than the 70K repository used for the IQS. With a larger repository file, the matcher is more likely to report high-scoring false matches. There is a chance that some of these false matches will score higher than the true mate, pushing the true mate further down the candidate list. An example of the process is shown in Figure 17. Note that the score remains constant as the database size increases, the ranking of candidates changes in relation to the other candidates in the database. For this reason, performance projections must be based on matcher scores, not on rank.

Small Database

Rank	ID	Score
1	M-1	4300
2	NM-1	4210
3	NM-2	3950
4	NM-3	3930
5	NM-4	3920
6	NM-5	3910
7	NM-6	3900
8	NM-7	3890

Large Database

Rank	ID	Score
1	NM-10	4500
2	NM-11	4350
3	NM-12	4330
4	NM-13	4325
5	M-1	4300
6	NM-14	4290
7	NM-1	4210
8	NM-2	3950
9	NM-3	3930
10	NM-4	3920

Key
- Mate
- Existing Non-Mates
- New Non-mates

Figure 17. Impact of Repository Size on Rank

Projecting operational performance to the AFIS repository requires adjusting performance measurements made on the ATB to incorporate effects of the larger repository size. It is not possible using the 70K repository available for the IQS study to project FARs of less than 10^{-7} to 10^{-9}. To maintain selectivity within reasonable bounds on the IAFIS 40 million repository, a FAR on the order of 10^{-10} to 10^{-11} is needed. Projection of FAR to the larger repository size involves extrapolating the selectivity at various threshold scores for the new file size and measuring the associated reliability at these points. Two methods were used to project ATB performance results to an operational database size:

- The extreme value statistics approach is a standard method used by statisticians for predicting occurrences of rarely seen phenomena. It uses the sparse data associated with the maximum non-mate fingerprint score to estimate reliability and selectivity in the larger CMF. The extreme value approach has a smaller margin of error in its estimates than the traditional approach. [Kinnison, p. 55] LMC used the extreme value approach to develop a projection model that was included in the ATB. The model estimates the selectivity and reliability at various threshold scores. The LMC extreme value projection does not account for any adjustments as a result of image quality.
- A traditional statistics approach uses an extrapolation of the central distribution to estimate reliability and selectivity in the larger CMF and to corroborate the extreme value method. Mitretek developed a traditional statistical approach projection method to corroborate the extreme value statistics model.

Comparison of the traditional approach model and the LMC extreme value model for the Latent ATB showed comparable results.

6.2 Projections for the Latent System

Since only a limited number of high-scoring false matches were available from any of the individual data sets, the false match distributions for DS1, DS2, and DS3[12] were combined to create a combined distribution of 8,658 false matches, which has a lower variance than any of the individual data sets.

Figure 18 shows the tradeoffs between reliability and selectivity. The chart shows that at the desired 0.01 selectivity operating point,[13] the three projections estimate reliabilities between 92 and 95 percent. When adjusted for image quality, a 95 percent confidence interval for reliability is expected to be in the range 85 to 90 percent.

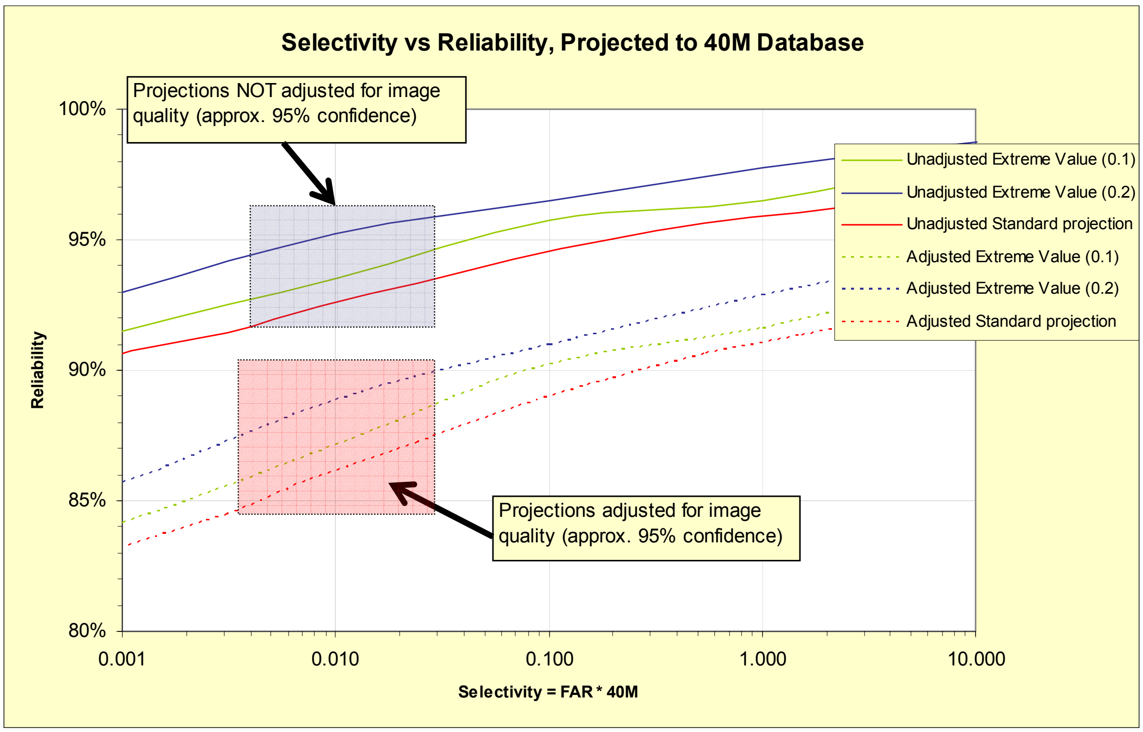


Figure 18. Latent System Reliability versus Selectivity Projected to CMF File Size (40M)

[12] Although the reliability results from DS3 were not useful, the false match distribution was not found to be biased and therefore was used to lower the estimates' variance.

[13] 0.01 selectivity @ 40M = 2.5 x 10^{-10} FAR.

6.3 Projections for the Ten-print System

The ten-print system projections show that this system (without modifications) will not perform as well as the latent system with two flat livescan images. The details are documented below.

Figure 19 shows the tradeoffs between reliability and selectivity. The chart shows that at the desired 0.01 selectivity operating point, the projected reliability for the ten-print algorithm suite is about 83 percent. When adjusted for image quality, the reliability may drop to about 74 percent.

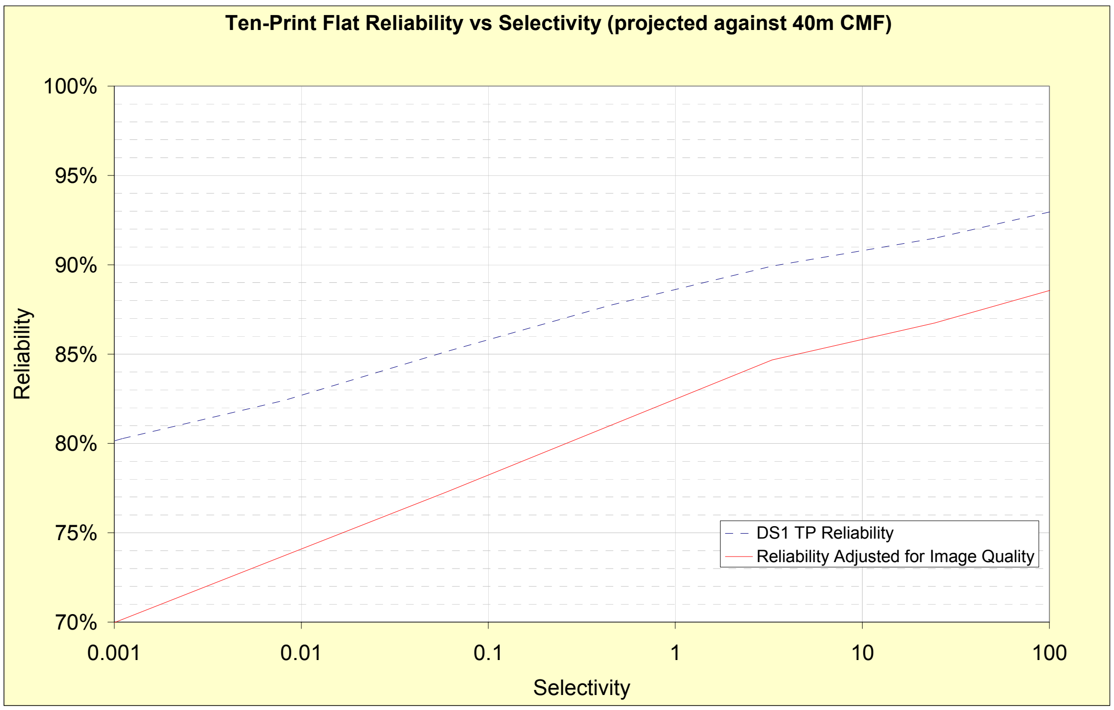


Figure 19. Ten-print System Reliability versus Selectivity Projected to CMF File Size (40M)

6.4 Estimating Reliability and Selectivity for Multi-Finger Data

Section 4.3 provides details of the performance of multi-finger slap and rolled searches. The slap results were based on a small number of searches, which did not provide enough false matches to clearly establish the relationship between FAR and reliability.

Although the slap test results are not adequate for performance projections, it is possible to make rough estimates of how multi-finger searches would perform. The IQS outlined a model of how multi-finger search performance could be estimated given the finding that the search performance of individual fingers is close to statistically independent. Therefore, in practice, the FRR for each two-finger search was very close to the product of the FRRs for separate left- and right-finger searches while keeping FAR constant. For example, in Table 16, the measured FRR for two fingers is very close to the estimate made using this method.

Table 16. Measured and Estimated FRR for Multiple Finger Searches (DS1)

	Measured	Estimated
Right Index	31.0%	
Left Index	33.0%	
2 fingers	12.0%	*10.2%*
4 fingers		*1.4%*
6 fingers		*0.2%*
(ten-print algorithms, 10^{-7} FAR, not adjusted for image quality)		

This model uses measured reliabilities at fixed FAR. The FAR did not vary when the model was tested by comparing estimated two-finger reliability to the measured value. The effect of this model when projecting to four or more fingers has not yet been determined; the degree of statistical independence may drop, in which case the estimates may represent the upper bounds on performance. Further analysis and testing with multi-finger data will be required.

The implications of fingerprint quality on multi-finger performance cannot be determined without testing with representative multi-finger data. Clearly, fingerprint image quality will have no more of an impact on multi-finger searches than two-finger searches, but whether image quality concerns are dramatically lessened cannot be determined without more testing using representative multi-finger data.

One very rough projection can be made from the slaps results shown in Table 9 and Table 11 (on pages 34 and 35). As a worst case hypothesis, FAR for 6 finger slap searches could be as high as 10^{-8} at a poor match threshold (scores > 3,200). This is fairly pessimistic. In ***every*** search conducted in the IQS and this study, the non-zero FAR as measured at a threshold of 7,000 is ***always*** less than 1/100th of the FAR as measured at a threshold of 3,200. This rough projection would suggest that a (fairly pessimistic) worst-case estimate of 6-finger slap selectivity would be less than 10^{-10} at a match score threshold of 7,000. The corresponding reliability at that point (from the previous table) is 92.9 percent.

The similarity of the FAR numbers for the slaps and BDM rolled data reinforces this projection; it also suggests that the starting point for the hypothesis is very pessimistic,

since the FAR of rolled 6 fingers at a 3,200 threshold is less than BDM's measurable limit of 4×10^{-9} (compared with the hypothesis starting point of 10^{-8}).

Therefore, if the slaps data set is representative, then without improving the segmentation algorithm and changing the ten-print algorithm suite, a rough but pessimistic projection of 6-finger performance against the CMF would be better than 92 percent reliability at a FAR of 10^{-10}. This is a rudimentary—but encouraging—projection.

Using the current ten-print algorithm search process as configured on the operational IAFIS is ***very likely*** to produce reliability better than 90 to 95 percent and selectivity less than 0.01 with 6 or 8 fingers segmented from slaps. Searching with 4 fingers is likely to result in worse reliability and selectivity; searching with 2 fingers is unlikely to be acceptable.

More testing with greater numbers of subjects is needed to determine whether IAFIS performance as tested by this small test data set is a valid and representative measure of AFIS performance.

6.5 System Resource Estimates

Since the existing IAFIS ten-print algorithm clearly cannot meet reasonable system requirements for two-print processing and the existing latent system has great processing demands, it would seem logical to consider combining the best aspects of each system. A limited analysis of the performance of the false matches observed in the latent system was performed as part of the IQS, but it was too late to include in the IQS report. This analysis indicated that it may be practical to combine portions of the ten-print matcher with the latent system so that accuracy could be improved while dramatically decreasing the processing requirements of the existing latent system.

The 303 highest scoring non-mates in the latent tests were analyzed separately to see whether the ten-print SSP or prescreener would have eliminated them. The results, though preliminary, are encouraging.

Of the high-scoring latent false matches, 9.6 percent would not have passed the ten-print SSP stage. This means that an IAFIS-based system designed to search flat fingerprints could use the ten-print SSP in combination with the latent system. This would reduce the latent systems' FAR with negligible increase in FRR while taking advantage of the efficiency of the SSP Filter.

SSP is not the only component of the ten-print system that could be used in combination with the latent system. The ten-print system's prescreener could be used at a relaxed threshold to reduce the number of non-mates sent to the latent system's matchers. Of the high-scoring latent false matches:

- 17.2 percent would not have passed the ten-print prescreen matcher at a CAXI threshold of 10 percent
- 23.4 percent would not have passed the ten-print prescreen matcher at a CAXI threshold of 5 percent.

The CAXI prescreen matcher is used very differently in the ten-print and latent systems. If the ten-print prescreen matcher (at a CAXI threshold of 10 percent) were used in

combination with the latent system, the latent system FAR would drop slightly, FRR would increase by 1 or 2 percent (see Table 3), but the processing requirements on the later matcher stages would be reduced by 90 percent.

As part of the E/SDS, Mitretek developed a resource model for a generalized IDENT/IAFIS architecture. The model estimates the computational resources that are required for different types of fingerprint searches. The model estimates the computational requirements for each stage of the search process by scaling the computational resources for each stage by the number of transactions to be processed by each stage.

The estimates presented are based on using untuned IAFIS algorithms. Since tuning of these algorithms would certainly reduce the resources required, these estimates are a worst-case estimate. Figure 20 shows the relative computational requirements as a function of the number of fingers used for the search process. These are rough estimates, making no adjustments for future technology, and assuming scalability. However, the model can be used to illustrate the relative benefit of using additional fingers in the search process.

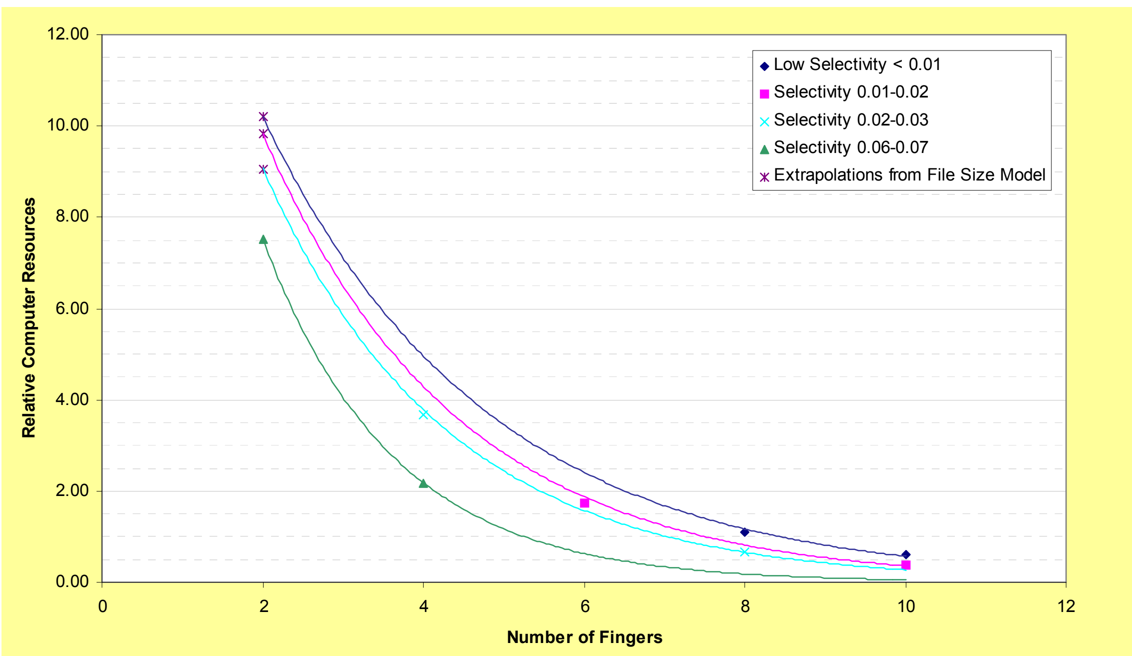


Figure 20. Relative Computer Resources versus Number of Fingers

The relative computational requirements as shown in Figure 20 clearly demonstrate the large benefit to be gained when more than 2 fingers are used for the search process. By going from 2 to 4 fingers, there is a drop in hardware cost of over 50 percent. Since hardware cost will be a large factor in the cost of the system, the cost implications of using 2 fingers for the search process will be a major factor in the selection of the final system architecture.

The general relationship between number of fingers and efficient use of resources is not specific to IAFIS: any well-designed AFIS will be able to be more efficient as fingers are added.

Section 7: Fingerprint Quality by Gender

No analyses of fingerprint quality and performance by gender were conducted during the original IQS, due to the rushed schedule of the study; the schedule of the IQS made it impossible to process all of the raw data that was collected.

DS2 included the gender of each subject, and an analysis of fingerprint quality by gender is revealing. DS2 is composed of 2,589 unmated subjects from the INS Recidivist file, including 2,227 males and 362 females. DS1, which was used for all performance analysis, did not include gender information, so it was not possible to compare matcher performance by gender.

Gender differences in fingerprint quality are subtle but profound:

- Minutiae-based quality metrics have very similar distributions for males and females.
- Ridge flow and classification quality measures are very clearly worse for females.

These differences indicate that system reliability and throughput will degrade as the proportion of females increases.

7.1 Ridge Quality by Gender

Classification and topology-based matching require clearly defined fingerprint ridge structure. The Sagem Morpho Ridge Flow (Gabor) metric clearly shows a gender difference, as shown in Figure 21.

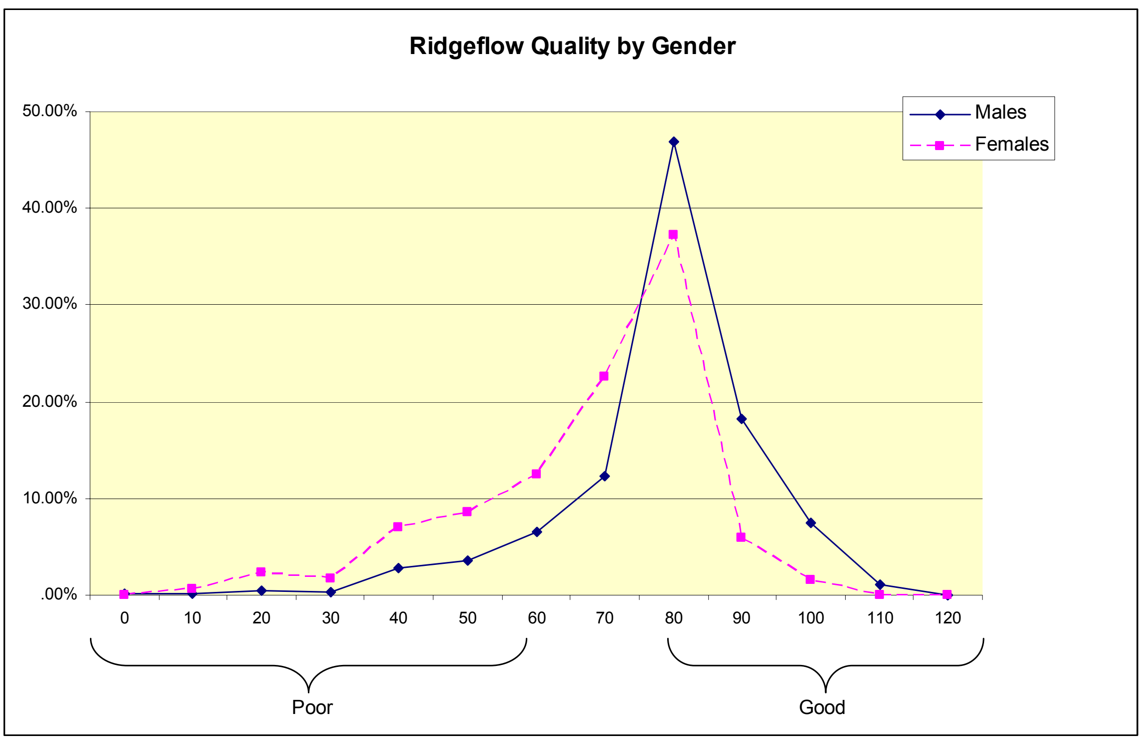


Figure 21. Ridge Flow Quality by Gender

Three characteristics of women's fingerprints make it difficult to determine ridge flow:

- The ridge diameter (and ridge frequency) is smaller than for men
- The ridges are shallower than for men
- Longitudinal (lengthwise) cracks are more common than for men

7.2 Classification and Filter Rate

As has been discussed elsewhere, the percentage of fingerprints that cannot be classified (fully referenced) has a substantial impact on system throughput and is directly related to filter rate.

Table 17 shows that about 85 percent of IDENT-quality female fingerprints cannot be classified, in contrast to about 58 percent of males. One way of interpreting these numbers is in terms of computing power—on average, matching female fingerprints will require about 150 percent of the processing needed to match male fingerprints.

Table 17. Unclassifiable Fingerprints by Gender

	% of Unclassifiable Fingerprints (Fully Referenced)
Males	57.9%
Females	84.8%
All	61.7%

7.3 Overall Fingerprint Quality by Gender

The overall fingerprint quality as measured by the UIQM is worse for females than for males, as should be expected since ridge flow is a component of UIQM. Figure 22 and Figure 23 show the corresponding quality distributions.

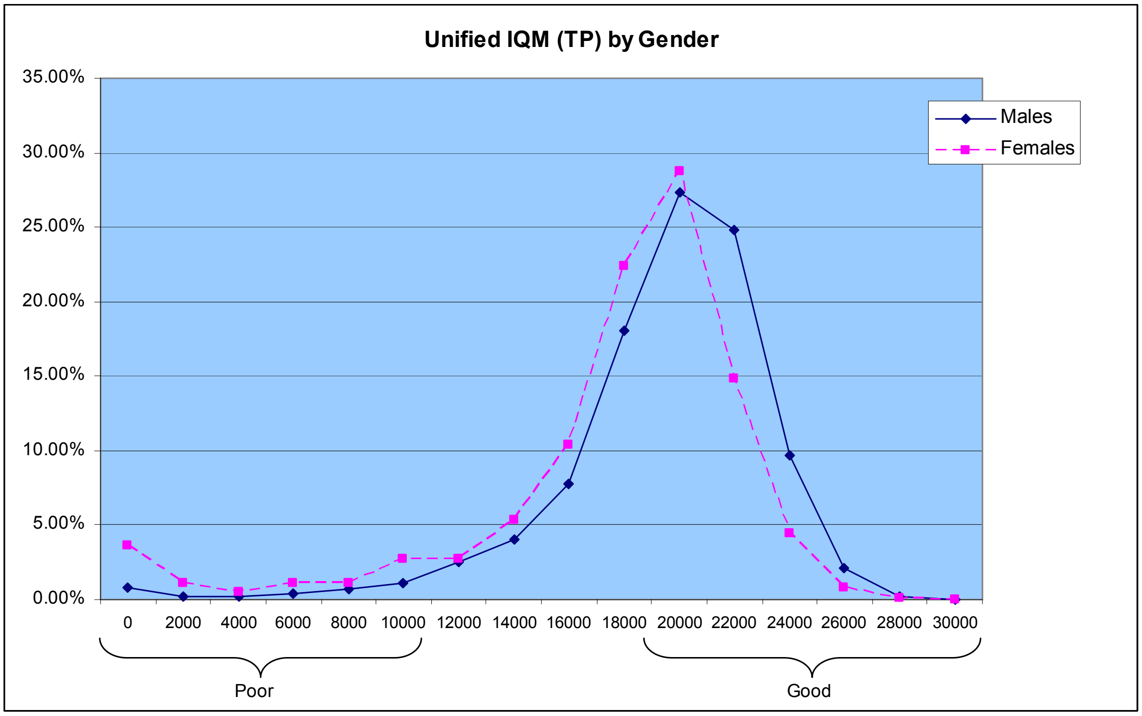


Figure 22. UIQM by Gender (Histogram)

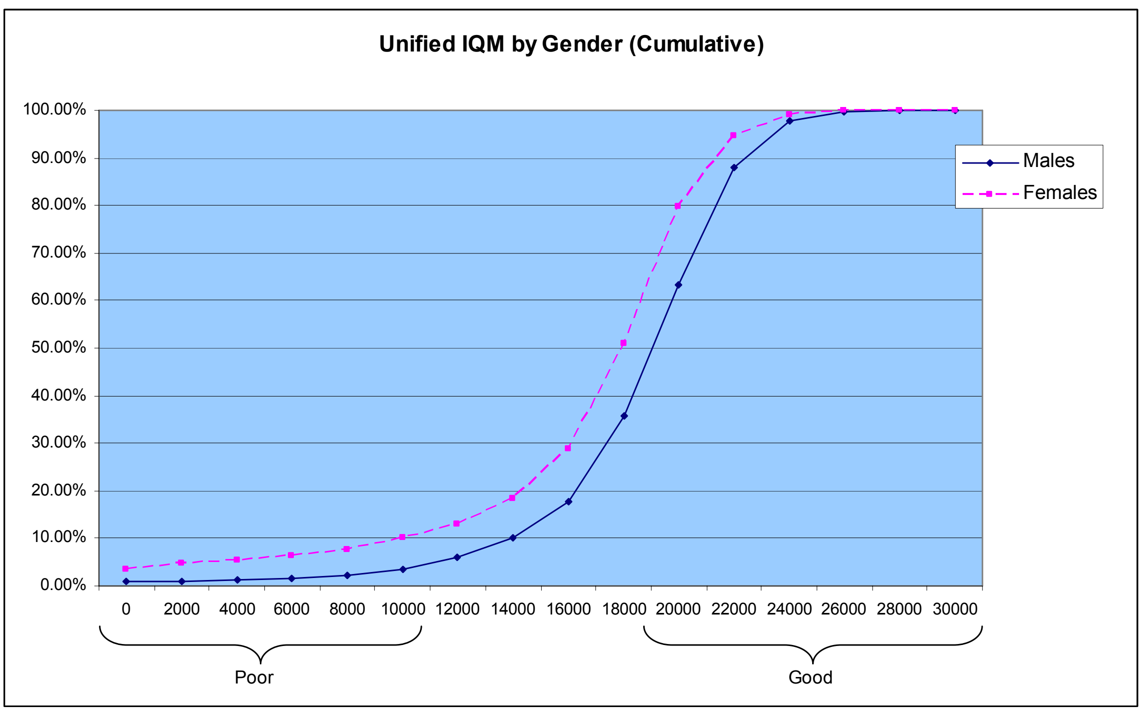


Figure 23. UIQM by Gender (Cumulative Distribution)

UIQM is not a perfect predictor of performance at medium or high scores, but poor UIQM scores are fairly effective predictors of failed matches. In Table 18, 6.7 percent of female

fingerprints are shown to be virtually unusable (in contrast to 2.4 percent of male fingerprints), and an additional 7 percent are very unlikely to match (in contrast to 5.4 percent of male fingerprints).

Table 18. Unified IQM by Gender

		Female	Male	Total
Very Poor	< 5,000	6.7%	2.4%	3.2%
Poor	5,000 - 11,000	7.0%	5.4%	5.7%
Medium	11,000 - 20,000	69.2%	31.0%	38.3%
Good	> 20,000	23.8%	63.6%	55.9%

7.4 Fingerprint Quality by Gender: Findings

Female fingerprints are poorer quality than male fingerprints. On average, matching female fingerprints will require about 150 percent of the processing needed to match male fingerprints. A greater proportion of female fingerprints are poor or very poor quality.

Clearly, performance and throughput will be engineering challenges for systems with large female populations. Analyses that predict the performance of a system should note what portion of the expected population is female. System designers should address whether minimum performance levels are expected or required to be the same for males and females.

Section 8: Slap Segmentation Accuracy

Two short studies were done of slap segmentation accuracy. Both studies used the slap segmentation software from Aware Corp. The initial study was done using Aware's "FingerPrint Image Segmenter" (Beta v. 12/01/2000), tested segmentation and matcher performance. The subsequent study used Aware's "Ten-Print Sequencing Library for Windows" (v 1.24, released June 21, 2001) only tested segmentation performance.

The data set used was the Civil 382 data set, which included subjects taken from civil IAFIS submissions in January and February 2000. No source, transaction-type, or scanner-type information was included with the data. This data set may or may not be representative, may be biased, and is large enough for general—but not definitive—results.

8.1 Segmentation Accuracy

During both studies, the images were segmented using the Aware software. During the second study, the segmentation of each image was visually inspected. During the earlier study, only samples were visually inspected.

For each image, the following types of problems were noted:

- Major problems
  - Unable to segment (fewer than four images per slap)
  - Unable to segment correctly (four images per slap, but fingers not correctly identified)
- Medium
  - Fingers correctly identified, but 1 or more fingers cropped substantially enough that matching would be affected
- Minor
  - Fingers correctly segmented, but minor cropping or rotation problems with 1 or more fingers (unlikely to have a dramatic affect on matching)

Determining when segmentation failed was a problem; in the "unable to segment" cases, the Aware software noted that fewer than 4 fingers were found, but the other problem cases could only be identified through visual inspection. Although the software returned a Segmentation Quality result, it did not provide a meaningful way to distinguish successful and unsuccessful segmentations. Figure 24 and Figure 25 show samples of segmentation problems.



Figure 24. Sample Missegmentation

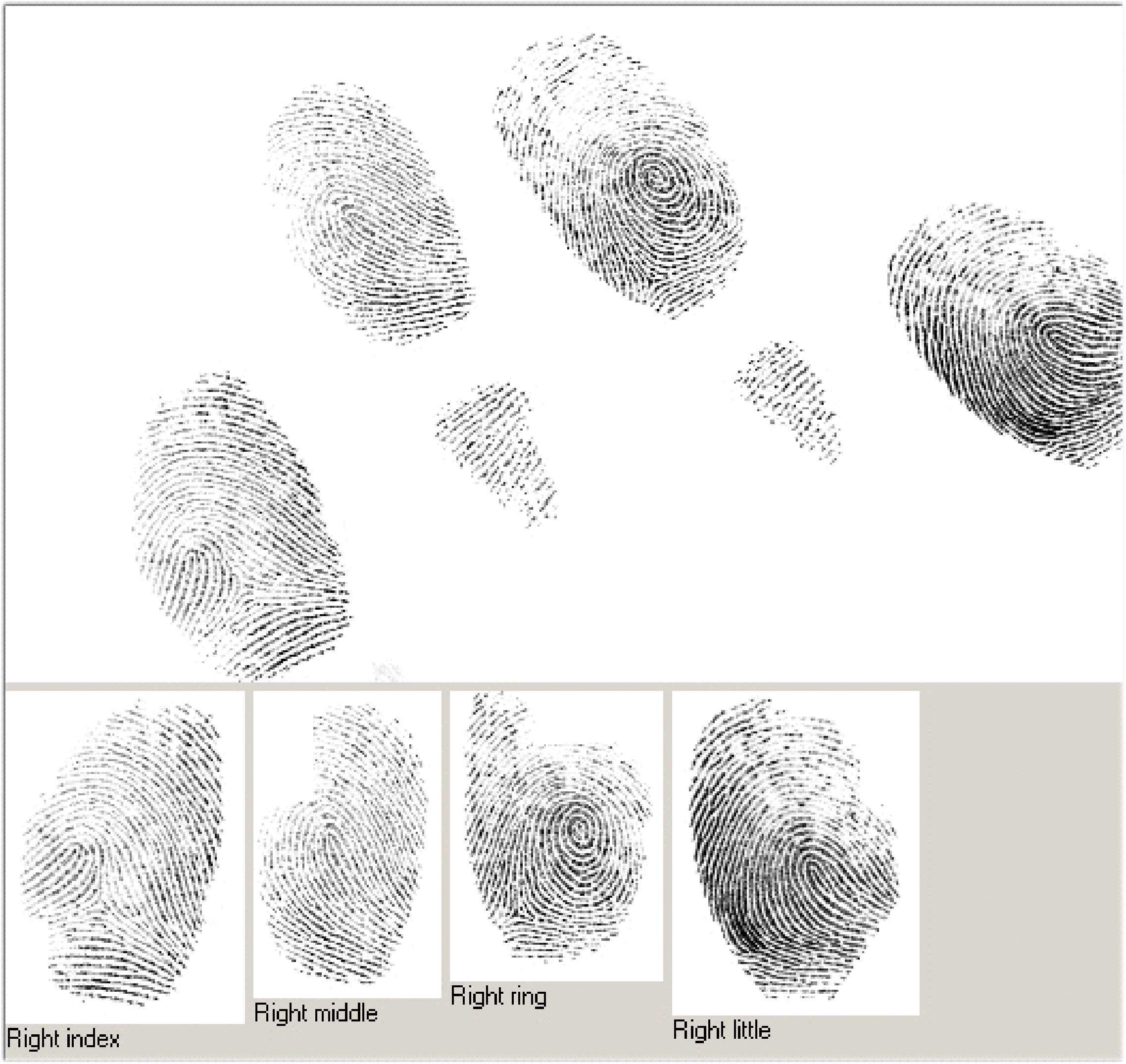


Figure 25. Sample of Minor Overcropping

Table 19 shows the segmentation accuracy during the second test. These results are much better than the results from the initial test. A comparison of the two tests is show in Table 20.

In the January 2001 study, only major problems were identified. In addition, the cases in which the segmenter incorrectly identified fingers were only visually inspected if matcher performance was affected, which almost certainly understated the number of problems. In that study, a number of preprocessing methods were used to attempt to successfully segment problem slaps.

Table 19. Segmentation Accuracy

		Image		Subject	
		Count	%	Count	%
Major	Unable to segment	4	0.5%	4	1.0%
	Missegmented	7	0.9%	7	1.8%
	Total	11	1.4%	11	2.9%
Medium	Image(s) overcropped substantially	20	2.6%	18	4.7%
Minor	Image(s) overcropped (minor)	25	3.3%	20	5.2%
	Incorrect rotation (minor)	2	0.3%	2	0.5%

Table 20. Segmentation Comparison

	Feb 2002 Tests				Jan 2001 (Primary Run)				Jan 2001 Never Segmented			
	Image		Subject		Image		Subject		Image		Subject	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Unable to segment	4	0.5%	4	1.0%	63	8.2%	54	14.1%	9	1.2%	9	2.4%
Missegmented	7	0.9%	7	1.8%	30[14]	3.9%	26	6.8%				
Total (Major)	11	1.4%	11	2.9%	93	12.2%	80	20.9%				

The "Never Segmented" results are the 9 images that previously would not segment under any circumstances. New performance is clearly better than this very generous comparison.

The segmenter attempts to find four fingerprints in the slap image, then reports the number of fingerprints detected, as well as bounding boxes for each image. This process is not as simple as it would first seem for several reasons:

- Successful searches can be conducted even if fewer than four fingerprints were detected.
- The segmenter may incorrectly segment four fingerprints.
- Successful searches can be conducted even if the segmenter incorrectly segmented the image.
- Different preprocessing parameters change the accuracy of the segmenter differently for each image.

14 Probably understated.

Note that the segmentability of the hands for a given subject is relatively independent; if the difficulty of segmenting the slaps for a subject were strongly correlated from right to left, the image and subject rates would be more similar.

8.2 Segmentation Findings

Converting a 4-finger image in to four upright fingerprint images can introduce errors, especially on poor-quality images. As discussed above, serious segmentation errors may occur in less than 2 percent of images. As shown in the tests discussed above, segmentation software has improved dramatically, and further improvements should be expected.

Segmentation software should be able to report in most cases whether segmentation was successful. This is an area needing improvement for the software used in this study.

IAFIS now uses LMC-developed segmentation software for sequence checking of IAFIS searches. The performance of the IAFIS segmentation software is reported to be superior to that of the Aware software, but the results of the comparison are not available at this time.

Section 9: Findings and Recommendations

9.1 Findings Relevant to Visa Processing

- ***Four or more flat fingerprints—preferably six or more—should be used when searching databases larger than 10 million subjects.***

  The IQS showed that IAFIS could not meet its FRR and FAR requirements using ten-print algorithms with two-finger searches of IDENT-quality data. IAFIS could meet its FRR and FAR requirements for two-finger searches using latent algorithms, but only at significantly increased processing cost. Using more fingers significantly reduces false reject rates (FRRs) and/or false accept rates (FARs), and will result in acceptable accuracy.

- ***Additional fingerprints significantly reduce processing requirements for searching large databases.***

  Using more fingers significantly improves processor performance. This improvement derives from the use of fingerprint classification indexing to reduce the number of candidates for each search. For each pair of fingers included in the search prints, the partitioning algorithm is able to cut the number of potential candidates approximately in half, which in turn halves processor requirements.

- ***The existing IAFIS algorithms could be reengineered to form a basis for improved flat fingerprint processing.***

  Portions of the current IAFIS ten-print and latent algorithms could be combined to produce an algorithm with flat fingerprint performance superior to the existing ten-print system and processing requirements significantly lower that the existing latent system.

- ***Female fingerprints are poorer quality than male fingerprints.***

  A greater proportion of female fingerprints are poor or very poor quality. On average, matching female fingerprints will require about 150 percent of the processing needed to match male fingerprints. Clearly, performance and throughput will be engineering challenges for systems with large female populations.

- ***Improvements in operational fingerprint quality will improve search accuracy.***

  The findings in the IQS are specific to the IDENT-quality data that was provided for this test. The preliminary slap fingerprint results indicate that just using the two index fingers from slap images would have substantially better performance accuracy than the two-finger IDENT data. This suggests that improvements in the operational quality of the data (such as by using more expensive fingerprint scanners) would improve performance accuracy over the IQS results.

- ***Operational fingerprint data will produce failure-to-enroll (FTE) errors.***

  Current IAFIS operations reject about 2.5 percent of civil submissions (rolled fingerprints) due to poor fingerprint quality. The quality of approximately 2 percent of INS IDENT flat fingerprints is so poor that it renders them virtually impossible to match using current IAFIS technology, and an additional 3 percent would be very unlikely to match.

  Slap fingerprints must be segmented into separate images; this process has an associated segmentation error rate that may result in FTE. The segmentation error rate using 2001 Aware automatic segmentation software is less than 2 percent of slap images (less than 3 percent of subjects).[15]

- ***Search fingerprint quality alone is an imperfect predictor of search performance.***

  A number of factors determine the accuracy of fingerprint matching:

  - Number of Fingers
  - Correspondence between Search and File images:
    - Overlapping areas
    - Lack of mutual distortion
  - Quality of both Search and File images:
    - Quality of ridge detail
    - Number of features
    - Size of image

  Image quality of the search fingerprint is only one of the factors determining the accuracy of the match.

- ***Poor search fingerprint quality is an effective predictor of search failure.***

  Fingerprints with poor quality are very unlikely to match. Effective minimum quality thresholds can be established using one or more IQMs.

- ***The methods by which sample data sets are collected can bias them so strongly that they are unusable for testing.***

  Great care must be taken in collecting the data sets used for testing. In particular, mated data sets should never be selected by using an AFIS; this process essentially filters out all of the hard-to-match fingerprints.

[15] Automated segmentation has improved dramatically since the IQS. The Aware software provides a data point for reference. It has not been compared with other segmentation software.

9.2 Recommendations for Visa Processing

- ***Slap fingerprints are appropriate for use in large-scale identification systems.***

  The significant improvement in FAR, FRR, and processing requirements as the number of search prints increases suggests that the use of slaps is the optimal compromise between matcher performance and operational constraints. The use of slaps offers operational improvements over the use of rolled fingerprints, since collecting slap fingerprints is a rapid process that does not require the same degree of operator training and "manhandling" of the subject. Operationally, collecting slaps and flat fingerprints is very similar. The use of slaps offers improvements in performance accuracy and efficiency over the use of flats.

- ***Large identification systems should be multimodal, incorporating demographic, facial, and possibly other biometric data.***

  The impact of FTE, FAR, and FRR errors arising from reliance on a single biometric can be largely overcome by incorporating alternative identifiers. An additional biometric would be particularly useful in processing subjects with poor quality fingerprints.

- ***Initiate a research program for ongoing analysis and comparison of emerging AFIS technology. Investigate the availability of new or improved algorithms, the possibility of improving existing algorithms, and the potential impacts of each.***

  Areas for further research include:

  - Slap segmentation
  - Pattern classification
  - Prescreen and detail matchers
  - Multimodal identification
  - New combinations of existing technology
  - Tuning for flat print searches
  - Tuning for flat print databases

- ***Collect representative test data sets for target search populations.***

  Current test data sets were drawn largely from criminal populations and may not be representative of visa applicants. Test sets representing children and the elderly are particularly needed.

- ***Develop and standardize policies and procedures to maintain operational quality.***

  Systems should be designed so that the equipment and operators are capable of collecting fingerprints of adequate quality, and ongoing measures (such as sampling or random tests) should be implemented to verify that the equipment and operators are, in fact, delivering fingerprints of adequate quality. Policies and procedures should address the following:

  - Training
  - User feedback
  - Scanner quality standards
  - Ongoing scanner quality tests, e.g. requiring two samples to be taken for each subject and a local match to verify that their quality is sufficient for subsequent use

- ***Design systems to measure ongoing operational performance.***

  Operational quality policies and procedures should be implemented in identification systems. Without auditing or sampling to determine operational FRR, FAR, and FTE rates, there is no means of determining ongoing system effectiveness. Lights-out systems should be instrumented to provide for audits. Template-only systems—those that do not store human-verifiable images that can be audited—have unknowable operational performance.

References

Kiebuzinski, George, Forman, Arthur V., O'Sullivan, James N., and Brown, David B., *Improving Fingerprint Submission Quality to Improve Reliability, Accelerate Throughput, and Reduce Cost*, unpublished draft.

Kinnison, Robert R., *Applied Extreme Value Statistics*, Battelle Press and MacMillan Publishing Company, New York, 1985, p. 55.

Lockheed Martin Corporation, *AFIS Technical Proposal*, GS000K94ALD0007, 21 August 1995.

Lockheed Martin Corporation, *AFIS/FBI Segment Design Document – Appendix B – AFIS Characterization and Search Algorithms*, 28 January 1998.

Lockheed Martin Corporation, *Algorithm Test Bed (ATB) User's Manual*, 31 August 2000.

Mitretek Systems (A. Hicklin), *Preliminary Analysis of Slap Fingerprint Performance,* Updated Version, February 22, 2002

Mitretek Systems (L. Nadel, et al), *IDENT/IAFIS Engineering/System Development Study, Image Quality Study Final Report*, December 7, 2000

U.S. Immigration and Naturalization Service (prepared for), *PMA-3 Benchmark Test Report*, October 7, 1998.

Appendix A: Sample Fingerprint Images from DS2

This appendix contains examples of the worst 0.1 percent, 2 percent, 5 percent, 15 percent, and best 0.1 percent of DS2 (Recidivist).

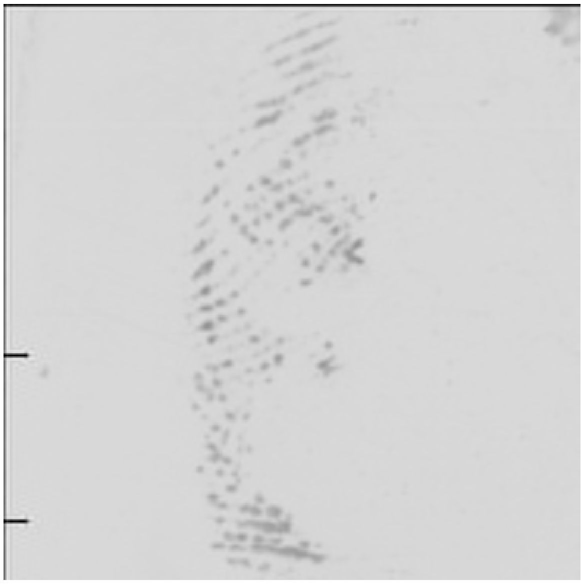


Figure 26. Example: Worst 0.1 percent of DS2

Finger	7
Position in File	0.1%
Sex	M
Minutiae	0
EqNum Minutiae	0
PattClass Quality	4
Unified Finger Quality TP	0



Figure 27. Example: Worst 2 percent of DS2

Finger	2
Position in File	2.0%
Sex	F
Minutiae	37
EqNum Minutiae	15
PattClass Quality	4
Unified Finger Quality TP	5099



Figure 28. Example: Worst 5 percent of DS2

Finger	7
Position in File	4.4%
Sex	F
Minutiae	39
EqNum Minutiae	19
PattClass Quality	4
Unified Finger Quality TP	9985



Figure 29. Example: Worst 15 percent of DS2

Finger	2
Position in File	14.2%
Sex	M
Minutiae	36
EqNum Minutiae	19
PattClass Quality	4
Unified Finger Quality TP	15009



Figure 30. Example: Best 0.1 percent of DS2

Finger	2
Position in File	100.0%
Sex	M
Minutiae	63
EqNum Minutiae	46
PattClass Quality	1
Unified Finger Quality TP	27614

Appendix B: Confidence Intervals

The following tables show the 95% confidence intervals and other details for selected tables and figures elsewhere in the document.

Table 21: Unclassifiable Fingerprints

	Dataset 1 Flat Index	Dataset 2 Flat Index	BDM Rolled Index	Slap Index	Slap Middle	Slap Ring	Slap Little
Subjects	3356	5178	7040	614	614	614	614
# Unclassified (=4)	973	3193	268	242	148	223	480
% Unclassified (=4)	29.0%	61.7%	3.8%	39.4%	24.1%	36.3%	78.2%
95% Confidence Interval (±)	1.5%	1.3%	0.4%	3.9%	3.4%	3.8%	3.3%

(see Figure 5. Unclassifiable Fingerprints)

Table 22: SSP Filter Rates by Finger Combinations (1 of 2)

	# Subjects	1 Finger index			2 Fingers index			4 Fingers index / middle		
BDM Rolled Data	3520	-			30.9%	±	0.4%	15.0%	±	0.3%
Slaps Data (Sample 1)	219	-			-			26.7%	±	2.5%
Slaps Data (Sample 2)	117	-			46.9%	±	4.0%	24.5%	±	3.4%
Flat Data (DS1)	3356	70.5%	±	0.8%	-			-		
Flat Data (DS1)	1678	-			44.6%	±	1.0%	-		
Flat Data (DS2)	2589	-			61.7%	±	0.9%	*30%* *Estimates from IQS*		

(see Table 2: Search Space Partitioning Filter Rates by Finger Combination)

Table 23: SSP Filter Rates by Finger Combinations (2 of 2)

	# Subjects	4 Fingers index / thumb			6 Fingers index / middle / ring			8 Fingers all except thumb			10 Fingers all		
BDM Rolled Data	3520	10.5%	±	0.3%	7.0%	±	0.2%	4.4%	±	0.2%	1.5%	±	0.1%
Slaps Data (Sample 1)	219	-			-			12.2%	±	1.7%	-		
Slaps Data (Sample 2)	117	-			13.2%	±	2.1%	10.1%	±	1.9%	-		
Flat Data (DS1)	3356	-			-			-			-		
Flat Data (DS1)	1678	-			-			-			-		
Flat Data (DS2)	2589	*21%*			*14%*			*9%*			*3%*		
		Estimates from IQS											

(see Table 2: Search Space Partitioning Filter Rates by Finger Combination)

Table 24. Multi-finger Reliability for Flat Data

	Flat DS1					
	2 fingers			1 finger		
Poor Match (Score > 3,200)	93.2%	±	1.2%	73.5%	±	1.5%
Marginal Match (Score > 5,000)	89.9%	±	1.4%	72.1%	±	1.5%
Good Match (Score > 10,000)	77.7%	±	2.0%	63.9%	±	1.6%
Certain Match (Score > 16,000)	57.0%	±	2.4%	52.8%	±	1.7%
	1678 subjects			3356 subjects		

(see Table 8. Observed Reliability by Finger Combination for Flat and Rolled Fingerprints)

Table 25. Multi-finger FAR for Flat Data

	Flat DS1					
	2 fingers			1 finger		
Poor Match (Score > 3,200)	3.0E-06	±	3.1E-07	6.4E-06	±	3.2E-07
Marginal Match (Score > 5,000)	7.7E-08	±	5.0E-08	2.0E-06	±	1.8E-07
Good Match (Score > 10,000)		<	2.5E-08	1.7E-08	±	1.7E-08
Certain Match (Score > 16,000)		<	2.5E-08		<	1.2E-08
Minimum measurable	8.5E-09			4.3E-09		
	1678 subjects			3356 subjects		

(see Table 10. Observed FAR by Finger Combination for Flat and Rolled Fingerprints)

Table 26. Multi-finger Reliability for Rolled Data

	BDM 3520 Rolled														
	All fingers			8 fingers			6 fingers			4 fingers			2 fingers		
Poor Match (Score > 3,200)	98.7%	±	0.5%	98.7%	±	0.5%	98.9%	±	0.5%	99.0%	±	0.5%	99.0%	±	0.5%
Marginal Match (Score > 5,000)	98.7%	±	0.5%	98.7%	±	0.5%	98.9%	±	0.5%	99.0%	±	0.5%	98.9%	±	0.5%
Good Match (Score > 10,000)	98.4%	±	0.6%	98.4%	±	0.6%	98.5%	±	0.6%	98.9%	±	0.5%	98.4%	±	0.6%
Certain Match (Score > 16,000)	97.9%	±	0.7%	97.9%	±	0.7%	98.2%	±	0.6%	98.2%	±	0.6%	97.2%	±	0.8%
	1825 subjects														

(see Table 8. Observed Reliability by Finger Combination for Flat and Rolled Fingerprints)

Table 27. Multi-finger FAR for Rolled Data

	BDM 3520 Rolled												
	All fingers		8 fingers		6 fingers		4 fingers			2 fingers			
Poor Match (Score > 3,200)	<	1.2E-08	<	1.2E-08	<	1.2E-08	4.8E-07	±	8.6E-08	1.3E-05	±	4.5E-07
Marginal Match (Score > 5,000)	<	1.2E-08	<	1.2E-08	<	1.2E-08		<	1.2E-08	7.9E-07	±	1.1E-07
Good Match (Score > 10,000)	<	1.2E-08	<	1.2E-08	<	1.2E-08		<	1.2E-08		<	1.2E-08
Certain Match (Score > 16,000)	<	1.2E-08	<	1.2E-08	<	1.2E-08		<	1.2E-08		<	1.2E-08
Minimum measurable	4.1E-09												
	3520 subjects												

(see Table 10. Observed FAR by Finger Combination for Flat and Rolled Fingerprints)

Table 28. Multi-finger Reliability for Slap Data (1 of 2)

	Segmentation Method 1											
	8 fingers			6 fingers			4 fingers			2 fingers		
Poor Match (Score > 3,200)	96.5%	±	3.4%	97.4%	±	2.9%	97.4%	±	2.9%	98.3%	±	2.4%
Marginal Match (Score > 5,000)	96.5%	±	3.4%	97.4%	±	2.9%	97.4%	±	2.9%	98.3%	±	2.4%
Good Match (Score > 10,000)	94.7%	±	4.1%	96.5%	±	3.3%	95.7%	±	3.7%	93.9%	±	4.4%
Certain Match (Score > 16,000)	91.2%	±	5.2%	92.2%	±	4.9%	93.0%	±	4.6%	87.0%	±	6.2%
# Subjects	113			115			115			115		
Note: Preliminary results												

(see Table 9. Observed Reliability by Finger Combination for Slap Fingerprints)

Table 29. Multi-finger Reliability for Slap Data (2 of 2)

	Segmentation Method 2					
	8 fingers			4 fingers		
Poor Match (Score > 3,200)		>	98.6%		>	98.6%
Marginal Match (Score > 5,000)		>	98.6%		>	98.6%
Good Match (Score > 10,000)	98.6%	±	1.55%	97.2%	±	2.2%
Certain Match (Score > 16,000)	95.4%	±	2.78%	95.0%	±	2.9%
# Subjects	218					
Note: Preliminary results						

(see Table 9. Observed Reliability by Finger Combination for Slap Fingerprints)

Table 30. Multi-finger FAR for Slap Data (1 of 2)

	Segmentation Method 1										
	8 fingers		6 fingers		4 fingers		2 fingers		
Poor Match (Score > 3,200)	<	3.6E-07	<	3.6E-07	<	3.6E-07	5.1E-06	±	1.6E-06
Marginal Match (Score > 5,000)	<	3.6E-07	<	3.6E-07	<	3.6E-07	2.4E-07	±	3.4E-07
Good Match (Score > 10,000)	<	3.6E-07	<	3.6E-07	<	3.6E-07		<	3.6E-07
Certain Match (Score > 16,000)	<	3.6E-07	<	3.6E-07	<	3.6E-07		<	3.6E-07
# Subjects	117								
Note: Preliminary results									

(see Table 11. Observed FAR by Finger Combination for Slap Fingerprints)

Table 31. Multi-finger FAR for Slap Data (2 of 2)

	Segmentation Method 2			
	8 fingers		4 fingers	
Poor Match (Score > 3,200)	<	3.6E-07	<	3.6E-07
Marginal Match (Score > 5,000)	<	3.6E-07	<	3.6E-07
Good Match (Score > 10,000)	<	3.6E-07	<	3.6E-07
Certain Match (Score > 16,000)	<	3.6E-07	<	3.6E-07
# Subjects	218			
Note: Preliminary results				

(see Table 11. Observed FAR by Finger Combination for Slap Fingerprints)

Glossary

AFIS	Automated Fingerprint Identification System – any automated system that can extract features from a fingerprint image and compare sets of these features for the purpose of subject identification; also, the fingerprint identification segment of IAFIS.
ATB	Algorithm Test Bed – Scaled-down AFIS used for testing; developed by Lockheed Martin Corporation
BDM 3520	Basic Demonstration Model Data Set – 3.520 ten-print rolled fingerprints used for testing.
Biometric Data	Information gathered from the physical features of a person that can be digitized and used for the purpose of identification. Current technology includes data extracted from fingers, palms, face, retina, etc.
Candidate	A potential match, retrieved from a repository and proposed by a comparison system, of the current criminal subject or applicant.
CAXI	Core and Axis Independent – A first stage matcher component employed by AFIS/FBI.
CMF	Criminal Master File – the IAFIS criminal ten-print file that contains fingerprint feature information. It is owned by the AFIS segment and is indexed by FBI Number (FNU).
COTS	Commercial Off-the-Shelf
CPU	Central Processing Unit
Demographic Data	Also known as biographic or descriptive data, information associated with an individual that may be described alphanumerically and used for identification purposes, e.g., color of eyes, date of birth, etc.
DOJ	Department of Justice
DOS	Department of State
EFTS	Electronic Fingerprint Transmission Specification – FBI document number CJIS-RS-0010 (V7), January 29, 1999. Describes the FBI's implementation of the national standard ANSI/NIST-ITL, 1-2000 Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information. (The latest version of the ANSI/NIST standard was approved on July 27, 2000 and is a consolidation of ANSI/NIST-CSL 1-1993 and ANSI/NIST-ITL 1a-1997.) This standard specifies a common format to be used to exchange fingerprint, facial, scar, mark, and tattoo identification data effectively across jurisdictional lines or between dissimilar systems made by difference manufacturers.
E/SDS	Engineering/System Development Study
FAR	False Accept Rate (see Section 2.1.2)

FBI	Federal Bureau of Investigation (DOJ)
Features	Any physical characteristics of fingerprint ridges (e.g., minutiae, curvature) that may by captured and represented numerically in order to define the fingerprint for the purpose of automated matching.
Fingerprint Examiner	An expert technician trained in the science of fingerprint identification.
Flat	A fingerprint impression made by pressing the finger to the medium or capture device without any rolling. Compared to a rolled impression, a flat is easier and faster to obtain, especially with an uncooperative subject. Without the finger edges, however, it may contain as little as 50 percent of the ridge and feature information of a rolled print and is thus much less valuable in latent print searching and comparison.
FP	Abbreviation for "fingerprint"
FRR	False Reject Rate (see Section 2.1.2)
IAFIS	Integrated Automated Fingerprint Identification System – the FBI's ten-print criminal history and latent fingerprint processing system. It receives requests, performs subject (name) search for ten-print requests, performs AFIS search against the Bureau's repositories, transmits a response to the originating agency, and performs appropriate file maintenance.
IDENT	Automated Biometric Identification System – currently, the primary alien identification system of INS Enforcement Branch, it is based on two fingerprints, a mugshot photo and demographic information. It searches a criminal alien file (LO) and a file of EWI recidivists (RC).
III	Interstate Identification Index – the subject search segment of IAFIS
IQM	Image Quality Metric
IQS	Image Quality Study
IT	Information Technology
Latent	A fingerprint impression, either partial or whole, that has been "lifted" from a crime scene; also, an identification search in which a latent impression is the search print.
Livescan	A process whereby fingerprints are captured directly on a reading device and simultaneously translated into electronic images.
LMC	Lockheed Martin Corporation
LO	INS Lookout Database
Lookout Database	Replaced in IDENT/IAFIS by New Lookout Database. Abbreviated LO, the current criminal history repository of the INS. For each subject it contains two rolled index fingerprints, a mugshot photo (as available), and demographic information, all extracted from a ten-print card. Criteria for enrollment provided in an INS memo of August 1998.

LT Latent

Minutiae A type of fingerprint feature that identifies each ridge ending and bifurcation (where a ridge divides) according to its x and y coordinates and the immediate angle of the ridge. A set of minutiae data (as many as possible – usually 50-150 for a rolled or flat impression, many fewer for a latent impression) may define a fingerprint for matching purposes.

NIST National Institute of Standards and Technology (formerly NBS, National Bureau of Standards).

One-to-One Verification In general, a comparison process whereby the identity of a known individual is confirmed. An identifier (FNU, FIN, etc.) and one or more subject fingerprints are supplied to the system, which then performs an automated comparison between the subject features and file features. Compare this to the typical fingerprint search, which is one-to-many.

Recidivist Database Abbreviated RC, the INS repository of aliens apprehended for EWI who have not otherwise violated U.S. law. For each subject it contains two flat index finger feature sets, a facial photo image, and biographical information, all captured by IDENT.

Roll(ed) A fingerprint impression made by rolling the finger and thus capturing the ridges at both sides of the finger. A rolled impression may include 50 percent more minutiae than the corresponding flat and is much preferred therefore for comparison with latent impressions.

Segment One of the three operational functions of IAFIS. These are:

ITN— Communications and workload management

III—Subject search

AFIS— Automated fingerprint matching

SSP Search Space Partition – the technique employed at the front-end of a fingerprint matcher to reduce the number of file prints that must be compared by the matchers by indexing the fingerprints based on pattern classification.

Submission A ten-print or latent search of IAFIS system files. A ten-print submission is associated with an arrest, requires the intervention of an FBI service provider to verify proposed candidates, and may result in the update of system files. Contrast with remote search. A latent submission is associated with a crime for which no arrest has been made.

Ten-print A 14-block fingerprint record consisting of ten rolled images and four flat images (one for each thumb, one for each four-finger group)

TP Ten-print

TRP The IAFIS Technology Refreshment Program

Two-print	A fingerprint record, such as used by the INS for illegal alien recidivists, that contains two flat fingerprint images captured on a single-finger livescan device. The fingerprint images captured are determined by organizational policy (by default, the two index fingerprint images).
UIQM	Unified Image Quality Metric
Verification	(1) The process by which a fingerprint examiner, observing a side-by-side display of corresponding finger images, determines if the identities of a search subject and proposed file candidate(s) are in fact the same. It is largely synonymous with FIC/VFIC. (2) The process of confirming that a person is who he claims to be by matching his biometric record against that of his claimed identity. Also known as one-to-one comparison. The term is also widely used to imply one-to-one verification.
Verifier	A fingerprint examiner assigned to a verification function.